

Staircases in \mathbb{Z}^2

Felix Breuer*

Frederik von Heymann†

June 5, 2009

Abstract

A staircase in this paper is the set of points in \mathbb{Z}^2 below a given rational line in the plane that have Manhattan Distance less than 1 to the line. Staircases are closely related to Beatty and Sturmian sequences of rational numbers. Connecting the geometry and the number theoretic concepts, we obtain three equivalent characterizations of Sturmian sequences of rational numbers, as well as a new proof of Barvinok’s Theorem in dimension two, a recursion formula for Dedekind-Carlitz polynomials and a partially new proof of White’s characterization of empty lattice tetrahedra. Our main tool is a recursive description of staircases in the spirit of the Euclidean Algorithm.

1 Introduction

Motivated by the study of lattice points inside polytopes, in this paper we seek to understand the set of lattice points “close” to a rational line in the plane. To this end we define a staircase in the plane to be the set of lattice points in the half-plane below a rational line that have Manhattan Distance less than 1 to the line. We prove several properties of these point sets, most importantly we show that they have a recursive structure that is reminiscent of the Euclidean Algorithm.

Not surprisingly, staircases are closely related to the Beatty and Sturmian sequences defined in number theory (see [Sto76, FMT78, PS90, O’B02]), i.e. to sequences of the form $(\lfloor \frac{b}{a}n \rfloor - \lfloor \frac{b}{a}(n-1) \rfloor)_n$ for $a, b \in \mathbb{N}$ with $\gcd(a, b) = 1$. We show several elementary properties of these sequences from a geometric point of view. To our knowledge such a geometric approach to these sequences is not available in the prior literature. Our observations lead to three characterizations of these sequences (Theorem 4.1). One of these is known (see [GLL78, Fra05]) while the other two seem to be new.

We conclude the paper by giving several applications of our findings. Firstly, we give a new proof of a theorem by Barvinok in dimension 2. Barvinok’s Theorem states that the generating function of the lattice points inside a rational simplicial cone can be written as a short rational function. While Barvinok uses a signed decomposition of the cone into unimodular cones to achieve this result, we partition the cone into sets that have a short representation.

Secondly, these ideas can also be used to give a recursion formula for Dedekind-Carlitz polynomials. These are polynomials of the form $\sum_{k=1}^{a-1} x^{k-1} y^{\lfloor \frac{b}{a}k \rfloor}$ or, equiv-

*Institut für Mathematik, Freie Universität Berlin, E-mail: felix.breuer@fu-berlin.de
Research supported by the Deutsche Forschungsgemeinschaft within the research training group ‘Methods for Discrete Structures’ (GRK 1408).

†Faculty Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, E-mail: F.J.vonHeymann@tudelft.nl
During the largest part of this research funded by the DFG Emmy Noether program (HA 4383/1) and the Freie Universität Berlin.

alently, generating functions of the lattice points inside the open fundamental parallelepipeds of cones of the form $\text{cone}\left(\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix}\right)$. Our recursion formula answers a question from [BHM08].

Finally, we simplify ideas from [Sca85] and [Rez06] to give a partially new proof of White's Theorem, which characterizes three-dimensional lattice simplices that contain no lattice points except their vertices.

This paper is organized as follows. In Section 2 we give definitions of and present elementary facts about the objects we study. While Section 2 is rather concise, we elaborate more in Section 3, where staircases are examined from a geometric point of view. Most importantly we explain the recursive structure of staircases in Lemma 3.9, 3.11 and 3.12. In Section 4 we motivate the three characterizations of Sturmian sequences before summarizing them in Theorem 4.1. Section 5 is devoted to the proof of this theorem. In Section 6 we apply Lemma 3.12 to give a new proof of Barvinok's Theorem in dimension 2 and in Section 7 we use Lemma 3.11 to give a recursion formula for Dedekind-Carlitz sums. We conclude the paper in Section 8 by giving a partially new proof of White's Theorem.

2 Staircases and Related Sequences

In this section we give the basic definitions we are going to work with. In particular we introduce staircases, which are the main geometric objects we will analyze. Then we will define some related sequences of integers and state basic facts about them and their connection to staircases. We elaborate on the geometric point of view and give additional examples in Section 3.

Before we introduce staircases, here are some preliminaries: For any real number $r \in \mathbb{R}$ we define the *integral part* $\lfloor r \rfloor := \max\{z \in \mathbb{Z} \mid z \leq r\}$ of r . The *fractional part* $\{r\}$ of r is then defined by $r = \lfloor r \rfloor + \{r\}$. Given $0 < a, b \in \mathbb{N}$ there exist unique integers $(b \text{ div } a)$ and $(b \bmod a)$ such that $b = (b \text{ div } a) \cdot a + (b \bmod a)$ and $0 \leq b \bmod a < a$. Using these two functions we can write $\lfloor \frac{b}{a} \rfloor = b \text{ div } a$ and $\{\frac{b}{a}\} = \frac{b \bmod a}{a}$. We are going to use these two notations interchangeably.

Given $A, B \subset \mathbb{R}^2$ and $v \in \mathbb{R}^2$, we define $A + B := \{a + b \mid a \in A, b \in B\}$ and $-A := \{-a \mid a \in A\}$ and we use the abbreviations $A - B := A + (-B)$ and $A + v = A + \{v\}$. We will refer to $A + B$ as the *Minkowski sum* of A and B . The difference of sets is denoted with $A \setminus B := \{a \in A \mid a \notin B\}$. A *lattice point* is an element of \mathbb{Z}^2 . An affine (linear) *lattice transformation* of the plane is an affine (linear) automorphism of the plane that maps \mathbb{Z}^2 bijectively onto \mathbb{Z}^2 . A vector $v \in \mathbb{Z}^d$ is *primitive* if $\gcd(v_1, \dots, v_d) = 1$ or, equivalently, if $\mathbb{Z}^d \cap \text{conv}(0, v) = \{0, v\}$.

Now, what is a staircase? Let L be an oriented rational line in the plane. Then L defines a positive half-space H . The task is to describe the lattice points in H that are close to L in the sense that they have distance < 1 to the line in the Manhattan metric. Equivalently we consider those points $x \in \mathbb{Z}^2 \cap H$ from which we can reach a point in the other half-space by a single horizontal or vertical step of unit length. Such a set of points we call a staircase. See Figure 1 for two examples. Note that it is sufficient to depict the staircase only under a primitive vector generating the line (in the first example the vector $(3, 8)$), as after that (and before that) the same pattern of points is repeated.

We do not require L to pass through the origin (or any other element of \mathbb{Z}^2), contrary to what Figure 1 might suggest. But we will see later that we can get all the information we want by looking only at lines through the origin. Also, without loss of generality we will restrict our attention to lines with positive slope,

as negative slopes will give us, up to mirror symmetry, the same sets.

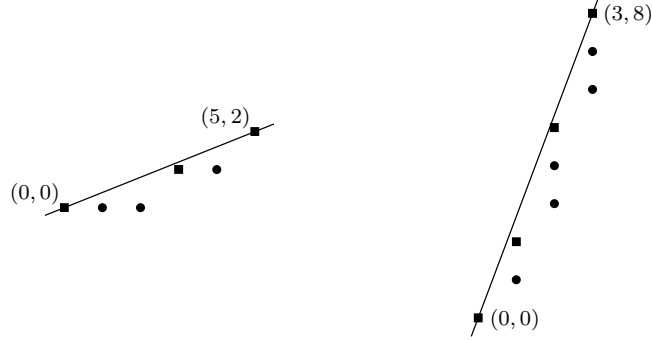


Figure 1: A part of the staircases $S_{5,2}$ and $S_{3,8}$. Corners are shown as boxes.

Let $0 < a, b \in \mathbb{N}$ with $\gcd(a, b) = 1$ and let $r \in \mathbb{R}$. These parameters define the line $L_{a,b,r} = \{x \in \mathbb{R}^2 \mid x_2 = \frac{b}{a}x_1 + r\}$. We denote the closed half-spaces below and above that line by $H_{a,b,r}^+$ and $H_{a,b,r}^-$, respectively. Formally we define for any $\sigma \in \{+1, -1\}$ the *half-space* $H_{a,b,r}^\sigma$ as

$$H_{a,b,r}^\sigma = \left\{ x \in \mathbb{R}^2 \mid 0 \leq \sigma \left(\frac{b}{a} x_1 - x_2 + r \right) \right\}.$$

Most of the time we will use $+$ to represent $+1$ and $-$ to represent -1 and write $\sigma \in \{+, -\}$ for short. Also if $\sigma = +$ and/or $r = 0$ we will omit these parameters and write $H_{a,b}$ for $H_{a,b,0}^+$, and similarly for the symbols introduced below. The case $\sigma = +$ and $r = 0$ is of the largest interest to us, as all other cases can be reduced to this one.

We will give a precise formulation and proof of this in Lemma 3.2. Although the proof does not require additional tools and could already be given here, we will pursue the connection between certain sequences and our point sets first, and postpone all observations that are purely concerned with the point sets to Section 3. So let's start defining these sets properly, to make it more clear what we are talking about.

The following definitions are illustrated in Figure 2. The lattice points in $H_{a,b,r}^\sigma$ that are at distance less than 1 from the line in vertical and horizontal direction, respectively, are

$$\begin{aligned} \mathcal{V}_{a,b,r}^\sigma &= \mathbb{Z}^2 \cap H_{a,b,r}^\sigma \setminus (H_{a,b,r}^\sigma - \sigma e_2) \\ &= \left\{ z \in \mathbb{Z}^2 \mid 0 \leq \sigma \left(\frac{b}{a} z_1 - z_2 + r \right) < 1 \right\} \\ \mathcal{H}_{a,b,r}^\sigma &= \mathbb{Z}^2 \cap H_{a,b,r}^\sigma \setminus (H_{a,b,r}^\sigma + \sigma e_1) \\ &= \left\{ z \in \mathbb{Z}^2 \mid 0 \leq \sigma \left(\frac{b}{a} z_1 - z_2 + r \right) < \frac{b}{a} \right\}. \end{aligned}$$

Using this notation we now define the *staircase* $S_{a,b,r}^\sigma$ to be the set of points that are at distance less than 1 from the line in horizontal *or* vertical direction, and we define the *corners* $C_{a,b,r}^\sigma$ to be the lattice points that are at distance less than 1 in horizontal *and* vertical direction:

$$\begin{aligned} S_{a,b,r}^\sigma &= \mathcal{V}_{a,b,r}^\sigma \cup \mathcal{H}_{a,b,r}^\sigma, \\ C_{a,b,r}^\sigma &= \mathcal{V}_{a,b,r}^\sigma \cap \mathcal{H}_{a,b,r}^\sigma. \end{aligned}$$

In other words

$$\begin{aligned} S_{a,b}^\sigma &= \{z \in \mathbb{Z}^2 \mid z \in H_{a,b}^\sigma \text{ but } z - \sigma e_1 \notin H_{a,b}^\sigma \text{ or } z + \sigma e_2 \notin H_{a,b}^\sigma\} \\ C_{a,b}^\sigma &= \{z \in \mathbb{Z}^2 \mid z \in H_{a,b}^\sigma \text{ but } z - \sigma e_1 \notin H_{a,b}^\sigma \text{ and } z + \sigma e_2 \notin H_{a,b}^\sigma\}. \end{aligned}$$

See Figure 1 and also Figure 2. Clearly $C_{a,b} \subset S_{a,b}$. For any set $A \subset \mathbb{Z}^2$ and any $x \in \mathbb{Z}$ we call the set $\text{col}_x(A) = \{(x, y) \in A\}$ a *column* of A and for $y \in \mathbb{Z}$ we call the set $\text{row}_y(A) = \{(x, y) \in A\}$ a *row* of A . For any $0 < a, b \in \mathbb{N}$, every row and every column of $S_{a,b}$ contains at least one point and every row and every column of $C_{a,b}$ contains at most one point.

The sequence $(|\text{col}_x(S_{a,b})|)_{x \in \mathbb{Z}}$ is called the *column sequence* of $S_{a,b}$, the sequence $(|\text{row}_y(C_{a,b})|)_{y \in \mathbb{Z}}$ is called the *row sequence* of $C_{a,b}$ and so on.

In the following we summarize some basic facts about staircases. We omit the proofs as they are easy enough to do and would slow the pace of this section without giving the reader further insights. The reader may find it instructive, however, to check the validity of these facts by looking at examples such as those given in the figures of this section.

Fact 1. For all $0 < a, b \in \mathbb{N}$ and $\sigma \in \{+, -\}$

$$\begin{aligned} a \geq b &\Leftrightarrow \mathcal{H}_{a,b}^\sigma \subset \mathcal{V}_{a,b}^\sigma \Leftrightarrow \forall x : |\text{col}_x(S_{a,b}^\sigma)| = 1 \Leftrightarrow \forall y : |\text{row}_y(C_{a,b}^\sigma)| = 1, \\ a \leq b &\Leftrightarrow \mathcal{H}_{a,b}^\sigma \supset \mathcal{V}_{a,b}^\sigma \Leftrightarrow \forall y : |\text{row}_y(S_{a,b}^\sigma)| = 1 \Leftrightarrow \forall x : |\text{col}_x(C_{a,b}^\sigma)| = 1. \end{aligned}$$

In the former case we call $S_{a,b}^\sigma$ *flat* and in the latter case we call $S_{a,b}^\sigma$ *steep*, see Figure 2. Note that this implies $S_{a,b}^\sigma = \mathcal{V}_{a,b}^\sigma$ for flat and $S_{a,b}^\sigma = \mathcal{H}_{a,b}^\sigma$ for steep staircases.

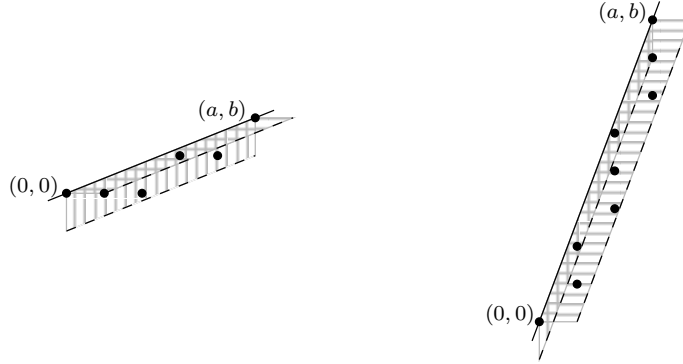


Figure 2: This figure shows the flat staircase $S_{5,2}$ and the steep staircase $S_{3,8}$ from Figure 1 together with parts of the corresponding sets $\mathcal{H}_{a,b}$ and $\mathcal{V}_{a,b}$. This illustrates Fact 1.

Fact 2. For all $n \in \mathbb{Z}$ the topmost point of $\text{col}_n(S_{a,b})$ is $(n, \lfloor \frac{b}{a}n \rfloor)$.

If $S_{a,b}$ is flat, we have seen in Fact 1 that for every $n \in \mathbb{Z}$ the set $\text{col}_n(S_{a,b})$ contains exactly one element, so all elements of $S_{a,b}$ have the form $(n, \lfloor \frac{b}{a}n \rfloor)$. If $S_{a,b}$ is steep, this is only true for the corners.

This description allows us to compute the difference in height between the topmost points in consecutive columns of $S_{a,b}$. For all $0 < a, b \in \mathbb{N}$ we define the sequence

$B_{a,b} = (B_{a,b}(n))_{n \in \mathbb{Z}}$ by

$$(2.1) \quad \begin{aligned} B_{a,b}(n) &:= \left\lfloor \frac{b}{a}n \right\rfloor - \left\lfloor \frac{b}{a}(n-1) \right\rfloor \\ &= \frac{b}{a} + \left\{ \frac{b}{a}(n-1) \right\} - \left\{ \frac{b}{a}n \right\}. \end{aligned}$$

Fact 3. If $a \leq b$ (i.e. $S_{a,b}$ is steep) then

$$|\text{col}_n(S_{a,b})| = B_{a,b}(n)$$

and if $b \leq a$ (i.e. $S_{a,b}$ is flat) then

$$|\text{col}_n(C_{a,b})| = B_{a,b}(n),$$

in particular $B_{a,b}$ is a 0, 1-sequence in this case.

The sequence $B_{a,b}$ is the key to connect the geometric description of “points close to a line” with notions from number-theory.

The sequences $(\lfloor \frac{b}{a}n \rfloor)_{n \in \mathbb{N}}$, $(B_{a,b}(n))_{n \in \mathbb{N}}$ and $(\underline{B}_{a,b}(n))_{n \in \mathbb{N}}$ (see below) are known as the *characteristic sequence*, the *Beatty sequence* and the *Sturmian sequence* of $\frac{b}{a}$, respectively.

These sequences are well studied in number theory, see [Sto76, FMT78, PS90, O'B02] for surveys, [Bro93] for historical remarks and [Fra05] for a discussion about the names of the sequences. However, only the characteristic sequences of irrational numbers are non-trivial from the point of view of number theory. They also appear in geometry, see Section 8.

For the rest of the section we will establish some definitions connected with the above sequences, and some basic properties of $B_{a,b}$, relating them to the staircases.

In this spirit, instead of working with $(\lfloor \frac{b}{a}n \rfloor)_{n \in \mathbb{N}}$, $(B_{a,b}(n))_{n \in \mathbb{N}}$ and $(\underline{B}_{a,b}(n))_{n \in \mathbb{N}}$, we will deal with $(\lfloor \frac{b}{a}n \rfloor)_{n \in \mathbb{Z}}$, $(B_{a,b}(n))_{n \in \mathbb{Z}}$ and $(\underline{B}_{a,b}(n))_{n \in \mathbb{Z}}$, respectively. This is due to the fact that we look at staircases of lines, not of rays.

Now let's define what a Sturmian sequence is. A sequence $s = (s_n)_{n \in \mathbb{Z}}$ of integers $s_n \in \mathbb{Z}$ is called *balanced* (at k) if $s_n \in \{k, k+1\}$ for all $n \in \mathbb{Z}$. If s is balanced at k , we can define a 0, 1-sequence $\underline{s} = (\underline{s}_n)_{n \in \mathbb{Z}}$, which we call the *reduced* sequence, by

$$\underline{s}_n = s_n - k.$$

Note that if $s = (c)_{n \in \mathbb{Z}}$ is constant, s is balanced at both c and $c-1$. In this case \underline{s} is defined with respect to c , i.e. \underline{s} is constant 0.

Lemma 2.1. $B_{a,b}$ is balanced at $\lfloor \frac{b}{a} \rfloor$. If $\frac{b}{a} \in \mathbb{Z}$, then $B_{a,b}(n) = \frac{b}{a}$ for all $n \in \mathbb{Z}$.

Proof. By (2.1) we know that $|B_{a,b}(n) - \frac{b}{a}| < 1$ and by definition $B_{a,b} \in \mathbb{Z}$. If $\frac{b}{a} \in \mathbb{Z}$, then the fractional parts in (2.1) are both 0, and thus the second statement is also true. \square

So Sturmian sequences $(\underline{B}_{a,b}(n))_{n \in \mathbb{N}}$ are well-defined. Furthermore, we now know that only two different integers appear in $B_{a,b}$, and that $\underline{B}_{a,b}$ tells us in which positions the larger integer of the two appears.

Given our geometric interpretation of $B_{a,b}$ from Fact 3 this means that a steep staircase has columns of only two different lengths and the reduced sequence

$(B_{a,b}(n))_{n \in \mathbb{N}}$ encodes which columns are long and which columns are short. We will return to the concept of reduction in Section 3.

To make the connection between the sequence $B_{a,b}$ and the point set $S_{a,b}$ more transparent, we introduce some more notation.

We will transfer the idea that all we want to know about the staircase can be found under a primitive integer vector in the line $L_{a,b}$ into the language of sequences.

For any sequence $s = (s_n)_{n \in \mathbb{Z}}$ we say that s is *periodic* with period $a \in \mathbb{N}$ if $(s_{n+a})_{n \in \mathbb{Z}} = (s_n)_{n \in \mathbb{Z}}$. We say that a is the *minimal period* of s if there is no period $a' \in \mathbb{N}$ of s with $a' < a$ and write $\mathbb{P}(s)$ for the minimal period of s . By (2.1), if $\gcd(a, b) = 1$, then $B_{a,b}$ is periodic with minimal period a .

For a periodic sequence s we define $\text{period}(s) = (s_n)_{0 \leq n < \mathbb{P}(s)}$. If s is a periodic 0,1-sequence, we write $\mathbf{1}(s)$ for the number of ones in $\text{period}(s)$. We will frequently represent s by the $\mathbb{P}(s)$ -tuple $\text{period}(s)$.

As $B_{a,b}$ describes the differences of the maximal heights in adjacent columns of $S_{a,b}$, these differences, accumulated between 0 and $a - 1$, must sum up to b .

We summarize the above observations into

Fact 4. If $0 < a, b \in \mathbb{N}$ and $\gcd(a, b) = 1$, then

$$\mathbb{P}(B_{a,b}) = a \quad \text{and} \quad \sum_{0 \leq n < a} B_{a,b}(n) = b.$$

In particular if $a > b$ (and thus $S_{a,b}$ is flat), then $\mathbf{1}(B_{a,b}) = b$.

To be more flexible when talking about parts of staircases respectively Beatty-sequences, we define the following. Given a sequence $s = (s_n)_{n \in \mathbb{Z}}$, a finite subsequence of the form

$$s|_{[x_0, x_1]} := (s_n)_{x_0 \leq n \leq x_1} \text{ for some } x_0 \leq x_1 \in \mathbb{Z}$$

will be called an *interval*. The number of elements $x_1 - x_0 + 1$ of $s|_{[x_0, x_1]}$ we will call the length of the interval and we will denote it by $\text{length}(s|_{[x_0, x_1]})$. If s is a 0,1-sequence, we will denote the number of ones in an interval $s|_{[x_0, x_1]}$ by $\text{ones}(s|_{[x_0, x_1]})$.

In Fact 4 we summed over the interval $B_{a,b}|_{[0, a-1]}$. But because of the periodicity of $S_{a,b}$ and $B_{a,b}$ we see that we could have used any interval of length $a - 1$. So for any fixed $i \in \mathbb{Z}$ the sequence $B_{a,b}(n + i)_{n \in \mathbb{Z}}$ also describes $S_{a,b}$. This gives rise to the following definition:

We say that sequences $s = (s_n)_{n \in \mathbb{Z}}$ and $s' = (s'_n)_{n \in \mathbb{Z}}$ are identical *up to shift* if there exists an $i \in \mathbb{Z}$ with $(s_{n+i})_{n \in \mathbb{Z}} = (s'_n)_{n \in \mathbb{Z}}$, in symbols $s \equiv s'$. Our goal in Section 4 will be to characterize Sturmian sequences up to shift.

3 Geometric Observations

In this section we develop some properties of staircases and their related sequences from a geometric point of view. The most important operation on staircases is for us the reduction, which we turn to in the latter half of this section. We start with some more elementary operations.

Throughout this section let $0 < a, b \in \mathbb{N}$ such that $\gcd(a, b) = 1$ and let $\sigma \in \{+, -\}$.

Elementary Properties of Staircases. As we have already mentioned (and used) before, all staircases with a given slope, regardless whether it's the one above or below the line, are translates of each other. Hence they yield the same step sequence up to shift. Before we finally prove this, we state an elementary lemma.

Lemma 3.1. *Let $r \in \mathbb{R}$. The line $L_{a,b,r}$ contains a lattice point if and only if $r = \frac{k}{a}$ for some $k \in \mathbb{Z}$.*

Proof. Without loss of generality we can assume $-1 < r \leq 0$. For any point $z \in \mathbb{Z}^2$ the vertical distance to the line $L_{a,b}$ is $\frac{b}{a}z_1 - z_2 = \frac{k}{a}$ for some $k \in \mathbb{Z}$. So if $r \neq \frac{k}{a}$ for any $k \in \mathbb{Z}$, then $L_{a,b,r}$ cannot contain a lattice point.

That $r = \frac{k}{a}$ is sufficient for the existence of a lattice point follows directly from the extended Euclidean Algorithm. It can also be shown with this geometric argument: $L_{a,b,r}$ can contain at most one lattice point z with $0 \leq z_1 < a$, for if there were two distinct lattice points with this property then $\gcd(a, b) \neq 1$. On the other hand $\mathcal{V}_{a,b} \cap ([0, a) \times \mathbb{R})$ contains exactly a lattice points, one in each column. Only the lines $L_{a,b,-\frac{k}{a}}$ with $0 \leq k \leq a-1$ can intersect $\mathcal{V}_{a,b}$. So each of them has to contain at least one lattice point. \square

Lemma 3.2. *For every $0 < a, b \in \mathbb{N}$ and $r \in \mathbb{R}$*

1. $S_{a,b,r} = S_{a,b} + v$ and $C_{a,b,r} = C_{a,b} + v$ for some $v \in \mathbb{Z}^2$ and
2. $S_{a,b}^- = S_{a,b} + v$ and $C_{a,b}^- = C_{a,b} + v$ for some $v \in \mathbb{Z}^2$.

Proof. 1. By Lemma 3.1, if $\frac{k}{a} \leq r < \frac{k+1}{a}$ then $S_{a,b,r} = S_{a,b,\frac{k}{a}}$ and $C_{a,b,r} = C_{a,b,\frac{k}{a}}$. Hence we can assume without loss of generality $r = \frac{k}{a}$, so the line $L_{a,b,r}$ contains a lattice point $v = (v_1, v_2)$ with $v_2 = \frac{b}{a}v_1 + r$. Then

$$\begin{aligned} z \in H_{a,b,r} - v &\Leftrightarrow z_2 + v_2 \leq \frac{b}{a}(z_1 + v_1) + r \\ &\Leftrightarrow z_2 \leq \frac{b}{a}z_1 \Leftrightarrow z \in H_{a,b}. \end{aligned}$$

This implies the first claim.

2. By Lemma 3.1 there is no lattice point v' with $\frac{a-1}{a} < \frac{b}{a}v'_1 - v'_2 < 1$, so

$$\mathcal{V}_{a,b}^- = \left\{ z \in \mathbb{Z}^2 \mid 0 \leq -\left(\frac{b}{a}z_1 - z_2\right) \leq \frac{a-1}{a} \right\}.$$

Also by Lemma 3.1, there exists a lattice point v with $\frac{b}{a}v_1 - v_2 = -\frac{a-1}{a}$ and for this point v

$$\begin{aligned} \mathcal{V}_{a,b}^- - v &= \left\{ z \in \mathbb{Z}^2 \mid 0 \leq -\left(\frac{b}{a}(z_1 + v_1) - (z_2 + v_2)\right) \leq \frac{a-1}{a} \right\} \\ &= \left\{ z \in \mathbb{Z}^2 \mid 0 \leq -\left(\frac{b}{a}z_1 - z_2 - \frac{a-1}{a}\right) \leq \frac{a-1}{a} \right\} \\ &= \left\{ z \in \mathbb{Z}^2 \mid \frac{a-1}{a} \geq \frac{b}{a}z_1 - z_2 \geq 0 \right\}. \end{aligned}$$

Applying the first observation again, we obtain

$$\mathcal{V}_{a,b}^- - v = \mathcal{V}_{a,b}^+.$$

A similar argument shows $\mathcal{H}_{a,b}^- - v = \mathcal{H}_{a,b}^+$ for a suitable v . Now, because of Fact 1, $S_{a,b}^\sigma = \mathcal{V}_{a,b}^\sigma$ and $C_{a,b}^\sigma = \mathcal{H}_{a,b}^\sigma$ or $S_{a,b}^\sigma = \mathcal{H}_{a,b}^\sigma$ and $C_{a,b}^\sigma = \mathcal{V}_{a,b}^\sigma$, depending on whether $a > b$ or $a < b$, where $\sigma \in \{+, -\}$. Therefore the above calculations imply 3.2.2. \square

The previous operations translated the staircases by an integral vector. Now we will introduce some other useful operations. We denote the reflection at the main diagonal by \searrow and the reflection at the origin by \bowtie , i.e. we define $\searrow(x, y) = (y, x)$ and $\bowtie(x, y) = (-x, -y)$. Note that both induce involutions, i.e. self-inverse bijective maps, on sets of lattice points; so we understand a set of lattice points if and only if we understand its reflection. The effect of these two reflections on staircases is illustrated with an example in Figures 3 and 4 and formalized in Lemmas 3.3 and 3.4.

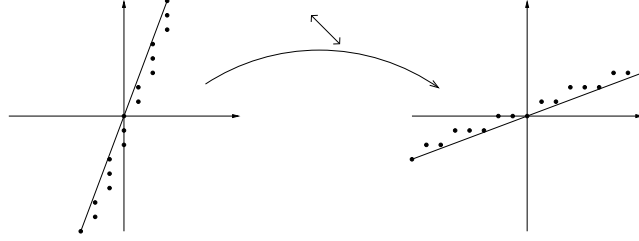


Figure 3: The reflection at the main diagonal swaps numerator and denominator of a staircase and places the points on the opposite side of the line. Here we see $\searrow S_{3,8} = S_{8,3}^-$.

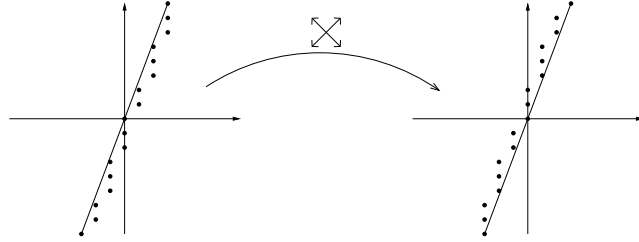


Figure 4: The reflection at the origin transforms a staircase below the line into a staircase above the line and vice versa. Here we see $\bowtie S_{3,8} = S_{3,8}^-$. Notice how the column sequence is reversed!

Lemma 3.3. $\searrow S_{a,b}^\sigma = S_{b,a}^{-\sigma}$ and $\searrow C_{a,b}^\sigma = C_{b,a}^{-\sigma}$.

In other words, reflection at the main diagonal swaps numerator and denominator of the slope and places the points on the opposite side of the line. See Figure 3.

Proof. We compute

$$\begin{aligned} (x, y) \in H_{a,b}^\sigma &\Leftrightarrow \sigma y \leq \sigma \left(\frac{b}{a}x\right) \Leftrightarrow \sigma x \geq \sigma \left(\frac{a}{b}y\right) \\ &\Leftrightarrow -(\sigma x) \leq -\sigma \left(\frac{a}{b}y\right) \Leftrightarrow (y, x) \in H_{b,a}^{-\sigma} \\ &\Leftrightarrow \searrow(x, y) \in H_{b,a}^{-\sigma}. \end{aligned}$$

This implies both

$$\begin{aligned} z - \sigma e_1 \notin H_{a,b}^\sigma &\Leftrightarrow \searrow(z) + (-\sigma)e_2 \notin H_{b,a}^{-\sigma} \quad \text{and} \\ z + \sigma e_2 \notin H_{a,b}^\sigma &\Leftrightarrow \searrow(z) - (-\sigma)e_1 \notin H_{b,a}^{-\sigma}. \end{aligned}$$

All three equivalences taken together give $\searrow S_{a,b}^\sigma = S_{b,a}^{-\sigma}$ and $\searrow C_{a,b}^\sigma = C_{b,a}^{-\sigma}$. \square

Lemma 3.4. $\bowtie S_{a,b}^\sigma = S_{a,b}^{-\sigma}$ and $\bowtie C_{a,b}^\sigma = C_{a,b}^{-\sigma}$.

Thus $(|\text{col}_n(S_{a,b}^\sigma)|)_n = (|\text{col}_{-n}(S_{a,b}^{-\sigma})|)_n$ and $(|\text{col}_n(C_{a,b}^\sigma)|)_n = (|\text{col}_{-n}(C_{a,b}^{-\sigma})|)_n$.

This means that reflection at the origin maps a staircase below the line to the staircase above the line and vice versa. This operation reverses the Beatty sequence of the staircase. See Figure 4.

Proof. We compute

$$\begin{aligned} (x, y) \in H_{a,b}^\sigma &\Leftrightarrow \sigma y \leq \sigma\left(\frac{b}{a}x\right) \Leftrightarrow -\sigma(-y) \leq -\sigma\left(\frac{b}{a}(-x)\right) \\ &\Leftrightarrow (-x, -y) \in H_{a,b}^{-\sigma} \Leftrightarrow \bowtie(x, y) \in H_{a,b}^{-\sigma}. \end{aligned}$$

which implies both $\bowtie S_{a,b}^\sigma = S_{a,b}^{-\sigma}$ and $\bowtie C_{a,b}^\sigma = C_{a,b}^{-\sigma}$, like in the proof of Lemma 3.3. For the second claim of the lemma we observe (using what we have already shown) that

$$\begin{aligned} (n, y) \in (\text{col}_n(S_{a,b}^\sigma)) &\Leftrightarrow (n, y) \in S_{a,b}^\sigma \Leftrightarrow (-n, -y) \in \bowtie S_{a,b}^\sigma \\ &\Leftrightarrow (-n, -y) \in S_{a,b}^{-\sigma} \Leftrightarrow (-n, -y) \in (\text{col}_{-n}(S_{a,b}^{-\sigma})). \end{aligned}$$

As this gives us for any fixed n a bijection between the sets $\text{col}_n(S_{a,b}^\sigma)$ and $\text{col}_{-n}(S_{a,b}^{-\sigma})$, their cardinality must be the same. The argument for $C_{a,b}^\sigma$ is analogous. \square

Putting Lemmas 3.2.2 and 3.4 together, we immediately obtain the non-obvious statement that reversing a Beatty sequence yields the same sequence up to shift.

Corollary 3.5. $(|\text{col}_n(S_{a,b}^\sigma)|)_n \equiv (|\text{col}_{-n}(S_{a,b}^\sigma)|)_n$.

Proof. $(|\text{col}_n(S_{a,b}^\sigma)|)_n \equiv (|\text{col}_n(S_{a,b}^{-\sigma}) + v|)_n \equiv (|\text{col}_n(S_{a,b}^{-\sigma})|)_n = (|\text{col}_{-n}(S_{a,b}^\sigma)|)_n$ \square

Similarly, Lemma 3.2.1 implies that $C_{a,b}$ and $C_{a,b,r}$ have the same column sequence for any r .

Recursive Description of Staircases. We now return to the operation called reduction, which we defined for balanced sequences in Section 2. First, let us observe the relation between Beatty and Sturmian sequences more closely. The following fundamental lemma tells us that, not surprisingly, Sturmian sequences are Beatty sequences with $a > b$ and vice versa.

Lemma 3.6. $\underline{B}_{a,b} = B_{a,b \bmod a}$. Conversely if s is a sequence balanced at $k \in \mathbb{N}$ and $\underline{s} = B_{a,b}$, then $s = B_{a,ak+b}$.

Proof. By (2.1) we observe that for any $k \in \mathbb{Z}$ such that both b and $b + ka$ are positive

$$B_{a,b}(n) + k = B_{a,b+ka}(n).$$

$B_{a,b}$ is balanced at $\lfloor \frac{b}{a} \rfloor = b \div a$ by Lemma 2.1. Note that by definition $b \bmod a = b - (b \div a)a$. So

$$\underline{B}_{a,b}(n) = B_{a,b}(n) - b \div a = B_{a,b \bmod a}(n).$$

Conversely if $\underline{s}(n) = B_{a,b}(n)$ and s is balanced at k , then

$$s(n) = \underline{s}(n) + k = B_{a,b}(n) + k = B_{a,b+ka}(n)$$

\square

How can this relation be phrased in terms of the staircases $S_{a,b}$ and $S_{a,b \bmod a}$? The following lemmas give an answer to this question. See $S_{5,13}$ and $S_{5,3}$ in Figure 5.

Lemma 3.7. *Let $0 < a < b$. The lattice transformation $A = \begin{pmatrix} 1 & 0 \\ b \operatorname{div} a & 1 \end{pmatrix}$ gives a bijection between $C_{a,b}$ and $S_{a,b \bmod a}$.*

The corners $C_{a,b}$ of the staircase $S_{a,b}$ are just the points of the smaller staircase $S_{a,b \bmod a}$ up to a lattice transform. Here “smaller” refers to both the number of lattice points in a given interval and the encoding length of the two parameters a and b . Note that the inverse of A is $A^{-1} = \begin{pmatrix} 1 & 0 \\ -b \operatorname{div} a & 1 \end{pmatrix}$.

Proof. As $S_{a,b}$ is steep, $\operatorname{col}_n(C_{a,b}) = \{(n, \lfloor \frac{b}{a}n \rfloor)\}$ and $\operatorname{col}_n(S_{a,b \bmod a}) = \{(n, \lfloor \frac{b \bmod a}{a}n \rfloor)\}$ by Facts 1 and 2. But

$$\begin{pmatrix} n \\ \lfloor \frac{b}{a}n \rfloor \end{pmatrix} = \begin{pmatrix} n \\ (b \operatorname{div} a)n + \lfloor \frac{b \bmod a}{a}n \rfloor \end{pmatrix} = A \begin{pmatrix} n \\ \lfloor \frac{b \bmod a}{a}n \rfloor \end{pmatrix}.$$

□

However, to obtain all points in $S_{a,b}$ from the corners $C_{a,b}$ we need to know which columns of $S_{a,b}$ are long and which are short (in the case $b > a$). It turns out that the corners in long columns are precisely the corners $C_{a,b \bmod a}$ of the smaller staircase, again up to the lattice transformation A .

Lemma 3.8. *Let $0 < a < b$. Then $\operatorname{col}_n(C_{a,b \bmod a})$ contains a point if and only if $\operatorname{col}_n(S_{a,b})$ is long.*

Proof. $S_{a,b \bmod a}$ is flat. So we know by Fact 3 that $(n, \lfloor \frac{b \bmod a}{a}n \rfloor) \in C_{a,b \bmod a}$ if and only if $1 = B_{a,b \bmod a}(n) = \underline{B}_{a,b}(n)$. But this just means that $\operatorname{col}_n(S_{a,b})$ is long. □

Taking the two lemmas together, we can describe every staircase $S_{a,b}$ in terms of the smaller corners and points in the smaller staircase $S_{a,b \bmod a}$. This result, together with our ability to swap the parameters a and b (by means of Lemmas 3.3 and 3.4) and the fact that the staircases $S_{a,1}$ are easy to describe, we obtain a recursive characterization of all staircases.

Let us look at an example, which is shown in Figure 5, before we formulate the recursion formally in Lemma 3.9. We want to express $S_{5,13}$ in terms of smaller staircases.

We know that the topmost points in each column are the points in $C_{5,13}$ and $C_{5,13}$ is just the image of $S_{5,3}$ under the lattice transformation $A = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$. Note that A keeps columns invariant.

We also know that $S_{5,13}$ has columns of lengths 2 and 3 and that the long columns are precisely those in which $C_{5,3}$ contains a point. So if we have an expression for $S_{5,3}$ and $C_{5,3}$, we can give an expression for $S_{5,13}$ and $C_{5,13}$.

To continue this argument inductively, we need to swap the parameters a and b , but this we can achieve by reflecting the staircases at the origin and at the main diagonal. So we reduce the problem of describing $S_{5,3}$ to the problem of describing $S_{3,5}$. We can now continue in this fashion, expressing $S_{3,5}$ in terms of $S_{3,2}$, in terms of $S_{2,3}$, in terms of $S_{2,1}$.

At this point we have finally reached a staircase with integral slope. These staircases have the nice property that all columns and all rows are identical and hence they can be described by a simple expression: the Minkowski sum of the lattice points on a line with those in an interval. This entire process is illustrated in Figure 5.

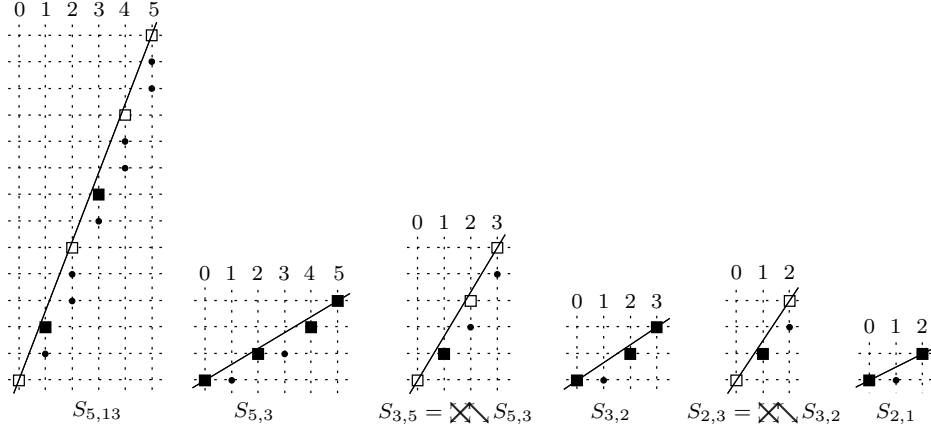


Figure 5: This figure show the recursive process of expressing $S_{5,13}$ in terms of smaller staircases, described in the text. In this figure, empty squares indicate the corners of long columns, filled squares corners of short columns. Note that the empty squares occur in $S_{a,b}$ precisely in the columns, in which there is an element of $C_{a,b \bmod a}$.

Lemma 3.9. Let $0 < a, b \in \mathbb{N}$ and $A = \begin{pmatrix} 1 & 0 \\ b \operatorname{div} a & 1 \end{pmatrix}$.

1. If $a < b$ and $\gcd(a, b) = 1$, then

$$\begin{aligned} C_{a,b} &= AS_{a,b \bmod a} \\ S_{a,b} &= AS_{a,b \bmod a} + \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ -(b \operatorname{div} a) + 1 \end{pmatrix} \right\} \\ &\quad \cup AC_{a,b \bmod a} + \left\{ \begin{pmatrix} 0 \\ -(b \operatorname{div} a) \end{pmatrix} \right\}. \end{aligned}$$

2. $C_{a,b} = \bowtie \searrow C_{b,a}$ and $S_{a,b} = \bowtie \searrow S_{b,a}$

3. If $b = 1$, then

$$\begin{aligned} C_{a,b} &= \left\{ \begin{pmatrix} ka \\ k \end{pmatrix} : k \in \mathbb{Z} \right\} \\ S_{a,b} &= \left\{ \begin{pmatrix} ka \\ k \end{pmatrix} : k \in \mathbb{Z} \right\} + \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} a-1 \\ 0 \end{pmatrix} \right\} \end{aligned}$$

In Section 4 we use this recursive structure to develop a characterization of Sturmian sequences. In Section 6 we employ the recursion to obtain short rational functions that enumerate the lattice points inside lattice polytopes in the plane, and in Section 7 for a representation of Dedekind-Carlitz polynomials that is computable in polynomial time.

Proof. 1. By Lemma 3.7 the first equation holds. Every column of $S_{a,b}$ contains a corner and every column contains at least $(b \operatorname{div} a)$ points. So $AS_{a,b \bmod a} + \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ -(b \operatorname{div} a) + 1 \end{pmatrix} \right\}$ contains all points in $S_{a,b}$ except the bottom-most points of the long columns. By Lemma 3.8 the long columns are precisely those in which $S_{a,b \bmod a}$ has a corner. So $AC_{a,b \bmod a} + \left\{ \begin{pmatrix} 0 \\ -(b \operatorname{div} a) \end{pmatrix} \right\}$ is precisely the set of bottom-most points of the long columns of $S_{a,b}$.

2. $S_{a,b} = \bowtie \searrow S_{a,b}^- = \bowtie \searrow S_{b,a}$ by Lemmas 3.3 and 3.4 and similarly for $C_{a,b}$.

3. If $b = 1$, then for all $k, n \in \mathbb{Z}$ we have $\lfloor \frac{b}{a} n \rfloor = k$ if and only if $ka \leq n \leq (k+1)a - 1$. Hence $\operatorname{row}_k(S_{a,b}) = \left\{ \begin{pmatrix} ka \\ k \end{pmatrix}, \dots, \begin{pmatrix} ka+a-1 \\ k \end{pmatrix} \right\}$ and $\operatorname{row}_k(C_{a,b}) = \left\{ \begin{pmatrix} ka \\ k \end{pmatrix} \right\}$. \square

Relation to the Euclidean Algorithm. This recursion is closely related to the Euclidean Algorithm, which takes as input two natural numbers $c_1, c_2 \in \mathbb{N}$. In each step $c_{i+1} = c_{i-1} \bmod c_i$ is computed. This continues until we reach a j such that $c_{j+1} = 0$ and $c_j \neq 0$. Then $c_j = \gcd(c_1, c_2)$.

Now suppose we want to determine $S_{b,a}$ and $C_{b,a}$ for some $b > a$. We flip the two parameters and then reduce the staircase, i.e. we apply 3.9.2 and 3.9.1. This reduces the problem to computing $S_{a,b \bmod a}$ and $C_{a,b \bmod a}$. Again we flip and reduce, which reduces the problem to computing $S_{b \bmod a, a \bmod (b \bmod a)}$ and $C_{b \bmod a, a \bmod (b \bmod a)}$ and we continue in this fashion. In other words, we put $c_1 = b$, $c_2 = a$ and $c_{i+1} = c_{i-1} \bmod c_i$ and compute the staircases $S_{c_i, c_{i+1}}$ and $C_{c_i, c_{i+1}}$ recursively, until we arrive at the case $S_{c_{j-1}, 1}$ and $C_{c_{j-1}, 1}$ which we can solve directly by 3.9.3. That we arrive in this case eventually follows by the correctness of the Euclidean Algorithm and the assumption that $\gcd(a, b) = 1$. Note also that this recursion terminates after few iterations. This is made precise in the following lemma.

Lemma 3.10. *Let $a, b \in \mathbb{N}$ and let $(c_n)_{n \in \mathbb{N}}$ denote the sequence defined by $c_1 = b$, $c_2 = a$ and $c_{i+2} = c_i \bmod c_{i+1}$. Then $\min \{j \in \mathbb{N} \mid c_{j+1} = 0\} \in \mathcal{O}(\log a)$.*

Proof. $(c_i)_{i \geq 2}$ is monotonously decreasing for all $a, b \in \mathbb{N}$, as by definition $c_{i+2} = c_i \bmod c_{i+1} < c_{i+1}$. Thus $c_i \operatorname{div} c_{i+1} \geq 1$ for $i \geq 2$ and so

$$c_i = \underbrace{(c_i \operatorname{div} c_{i+1})}_{\geq 1} c_{i+1} + \underbrace{(c_i \bmod c_{i+1})}_{= c_{i+2}} \geq \underbrace{c_{i+1} + c_{i+2}}_{\geq c_{i+2}} \geq 2c_{i+2}$$

for $i \geq 2$. Hence $c_{i+2k} \leq 2^{-k} c_i$, and so if $k \geq \log_2 c_i$, then $c_{i+2k} \leq 1$. In particular the minimal j such that $c_{j+1} = 0$ satisfies $j \leq 2 \log_2 c_2 + 2 \in \mathcal{O}(\log a)$. \square

Recursive Description of Parallelepipeds. Instead of describing the infinite set of lattice points in an entire staircase, one might want to describe finite subsets thereof, for example the set of lattice points in only “one period” of the staircase. We now give a recursion for the set of lattice points in the fundamental parallelepipeds of the cones $\operatorname{cone}\left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)$ and $\operatorname{cone}\left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right)$.

The **cone** generated by $v_1, \dots, v_n \in \mathbb{R}^d$ is the set

$$\operatorname{cone}(v_1, \dots, v_n) = \left\{ \sum_{i=1}^n \alpha_i v_i \mid 0 \leq \alpha_i \in \mathbb{R} \text{ for all } 1 \leq i \leq n \right\}.$$

A cone is rational if all the v_i are rational and it is simplicial if the v_i are linearly independent. The **fundamental parallelepiped** $\Pi_{\operatorname{cone}(v_1, \dots, v_n)}$ of a simplicial cone $\operatorname{cone}(v_1, \dots, v_n)$ is defined as

$$\Pi_{\operatorname{cone}(v_1, \dots, v_n)} := \left\{ \sum_{i=1}^n \alpha_i v_i \mid 0 \leq \alpha_i < 1 \text{ for all } 1 \leq i \leq n \right\}.$$

Note that any rational cone $\operatorname{cone}(v_1, \dots, v_n) \subseteq \mathbb{R}^m$ can be transformed unimodularly to a rational cone $\operatorname{cone}(\sigma e_j, v'_1, \dots, v'_n)$ with $\sigma \in \{+, -\}$ and $1 \leq j \leq m$. So we don't restrict ourselves by only looking at cones containing e_1 or $-e_2$ in the generators. With the above notation

$$\begin{aligned} \mathcal{V}_{a,b} \cap [0, a) \times \mathbb{R} &= \Pi_{\operatorname{cone}\left(\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix}\right)} \cap \mathbb{Z}^2 \\ \mathcal{H}_{a,b} \cap \mathbb{R} \times [0, b) &= \Pi_{\operatorname{cone}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix}\right)} \cap \mathbb{Z}^2, \end{aligned}$$

see Figure 6. This means that if $a < b$ (and hence $S_{a,b} = \mathcal{H}_{a,b}$), the points z in the staircase $S_{a,b}$ with $0 \leq z_2 < b$ are just the lattice points in the fundamental

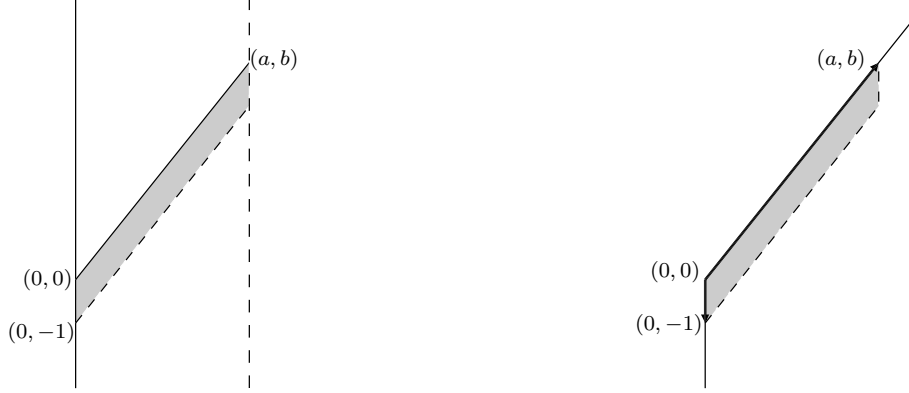


Figure 6: If we intersect $\mathcal{V}_{a,b}$ with $[0, a) \times \mathbb{R}$ we obtain the fundamental parallelepiped of the cone generated by $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} a \\ b \end{pmatrix}$.

parallelepiped $\Pi_{\text{cone}(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix})}$. The corners $z \in C_{a,b}$ with $0 \leq z_1 < a$ are just the lattice points in the fundamental parallelepiped $\Pi_{\text{cone}(\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix})}$.

To give an interpretation of our recursion in terms of fundamental parallelepipeds it is convenient to define the set $\Pi_{\text{cone}(v_1, \dots, v_n)}^\circ$ of lattice points (!) in the open fundamental parallelepiped of $\text{cone}(v_1, \dots, v_n)$ as

$$\Pi_{\text{cone}(v_1, \dots, v_n)}^\circ := \mathbb{Z}^2 \cap \left\{ \sum_{i=1}^n \alpha_i v_i \mid 0 < \alpha_i < 1 \right\}.$$

Note that if $n = 2$ and both v_1 and v_2 are primitive, then $\Pi_{\text{cone}(v_1, v_2)}^\circ \cup \{(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix})\} = \mathbb{Z}^2 \cap \Pi_{\text{cone}(v_1, v_2)}$. So it suffices to give a recursion for the sets of lattice points in open fundamental parallelepipeds.

We are going to use the following abbreviations:

$$\begin{aligned} \Pi_{\downarrow, a, b} &:= \Pi_{\text{cone}(\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix})} & \Pi_{\downarrow, a, b}^\circ &:= \Pi_{\text{cone}(\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix})}^\circ \\ \Pi_{\rightarrow, a, b} &:= \Pi_{\text{cone}(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix})} & \Pi_{\rightarrow, a, b}^\circ &:= \Pi_{\text{cone}(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix})}^\circ \end{aligned}$$

In terms of open parallelepipeds, Lemma 3.9 can now be phrased as follows. An example illustrating the somewhat involved expression in 3.11.1 is given in Figure 7.

Lemma 3.11. *Let $a, b \in \mathbb{N}$ and $A = \begin{pmatrix} 1 & 0 \\ b \operatorname{div} a & 1 \end{pmatrix}$.*

1. *If $0 < a < b$ and $\gcd(a, b) = 1$, then*

$$\begin{aligned} \Pi_{\downarrow, a, b}^\circ &= A \Pi_{\downarrow, a, b \operatorname{mod} a}^\circ \\ \Pi_{\rightarrow, a, b}^\circ &= A \Pi_{\downarrow, a, b \operatorname{mod} a}^\circ + \{(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}), \dots, (\begin{smallmatrix} 0 \\ -(b \operatorname{div} a) + 1 \end{smallmatrix})\} \\ &\quad \cup A \Pi_{\rightarrow, a, b \operatorname{mod} a}^\circ + (\begin{smallmatrix} 0 \\ -b \operatorname{div} a \end{smallmatrix}) \\ &\quad \cup (\begin{smallmatrix} a \\ b \end{smallmatrix}) + \{(\begin{smallmatrix} 0 \\ -1 \end{smallmatrix}), \dots, (\begin{smallmatrix} 0 \\ -b \operatorname{div} a \end{smallmatrix})\}. \end{aligned}$$

2. $\Pi_{\rightarrow, a, b}^\circ = \boxtimes \Pi_{\downarrow, b, a}^\circ + (\begin{smallmatrix} a \\ b \end{smallmatrix})$ and $\Pi_{\downarrow, a, b}^\circ = \boxtimes \Pi_{\rightarrow, b, a}^\circ + (\begin{smallmatrix} a \\ b \end{smallmatrix})$.

3. $\Pi_{\rightarrow, a, 1}^\circ = \emptyset$ and $\Pi_{\downarrow, a, 1}^\circ = \{(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}), \dots, (\begin{smallmatrix} a-1 \\ 0 \end{smallmatrix})\}$.

This allows us describe $\Pi_{\downarrow, a, b}^\circ$ in terms of $\Pi_{\downarrow, a, b \operatorname{mod} a}^\circ$ and $\Pi_{\rightarrow, a, b \operatorname{mod} a}^\circ$. The proof is similar to the one of Lemma 3.9 and we omit it for brevity.

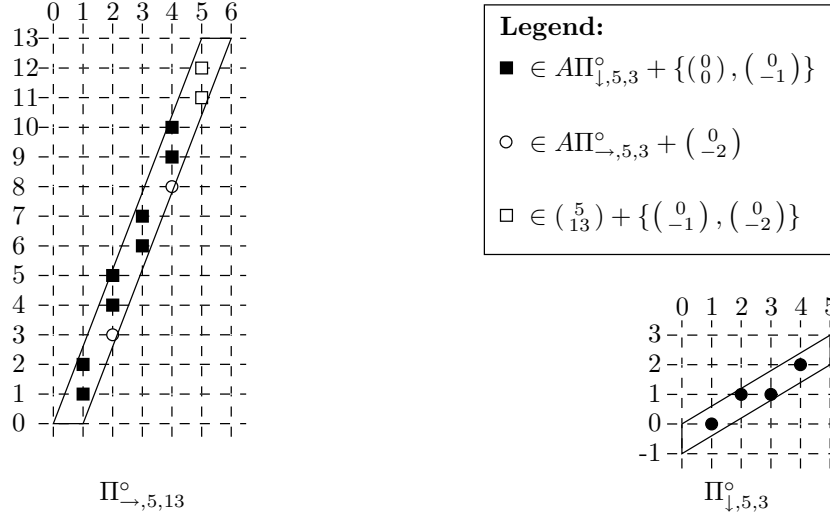


Figure 7: This figure illustrates the formula given in 3.11.1. $\Pi_{\rightarrow,5,13}^o$ is expressed in terms on $\Pi_{\downarrow,5,3}^o$ (shown) and $\Pi_{\rightarrow,5,3}^o$ (the corners of $\Pi_{\downarrow,5,3}^o$). The idea is the same as in Figure 5 and Lemma 3.9. However there is one important difference: Both $\text{col}_5(\Pi_{\downarrow,5,3}^o)$ and $\text{col}_5(\Pi_{\rightarrow,5,3}^o)$ are empty. These have to be added using the third term $\begin{pmatrix} 5 \\ 13 \end{pmatrix} + \left\{\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -2 \end{pmatrix}\right\}$.

Recursive Description of Triangles. We conclude this section by giving a similar recursion for triangles, see Figure 8. We write

$$\Delta_{a,b} := \text{conv} \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} a \\ 0 \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix} \right\}$$

and

$$\Delta'_{a,b} := \Delta_{a,b} \setminus \text{conv} \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix} \right\}$$

to denote closed and half-open triangles, respectively. The corresponding lattice point sets are denoted by $T_{a,b} := \Delta_{a,b} \cap \mathbb{Z}^2$ and $T'_{a,b} := \Delta'_{a,b} \cap \mathbb{Z}^2$.

The idea is now that for $0 < a < b$ the triangle $\Delta_{a,b}$ can be decomposed into two parts $\Delta'_{a,(b \bmod a)a}$ and $A\Delta_{a,b \bmod a}$. The former is defined by a line with integral slope and hence the set of lattice points $T'_{a,(b \bmod a)a}$ is easy to describe. The latter can be transformed into $\Delta_{b \bmod a,a}$ and we can obtain a description of the lattice point set $T_{b \bmod a,a}$ recursively. See Figure 8. The resulting recursion is given in Lemma 3.12 without proof.

Lemma 3.12. *Let $a, b \in \mathbb{N}$ and $A = \begin{pmatrix} 1 & 0 \\ b \bmod a & 1 \end{pmatrix}$.*

1. $T_{a,b} = AT_{a,b \bmod a} \cup T'_{a,(b \bmod a)a}$.
2. $T_{a,b} = \boxtimes \setminus \setminus T_{b,a} + \begin{pmatrix} a \\ b \end{pmatrix}$.
3. $T_{a,1} = \left\{ \begin{pmatrix} k \\ 0 \end{pmatrix} \mid 0 \leq k \leq a \right\} \cup \left\{ \begin{pmatrix} a \\ 1 \end{pmatrix} \right\}$.
4. *If $k \in \mathbb{N}$, then*

$$\begin{aligned}
T'_{a,ka} &= \bigcup_{0 < l \leq a} \left\{ \begin{pmatrix} l \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} l \\ lk-1 \end{pmatrix} \right\} \\
&= \left\{ \begin{pmatrix} 0 \\ m \end{pmatrix} \mid m \in \mathbb{N} \right\} + \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} a \\ 0 \end{pmatrix} \right\} \\
&\quad \setminus \left(\left\{ \begin{pmatrix} 0 \\ m \end{pmatrix} \mid m \in \mathbb{N} \right\} + \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ k \end{pmatrix}, \dots, \begin{pmatrix} a \\ ak \end{pmatrix} \right\} \right).
\end{aligned}$$

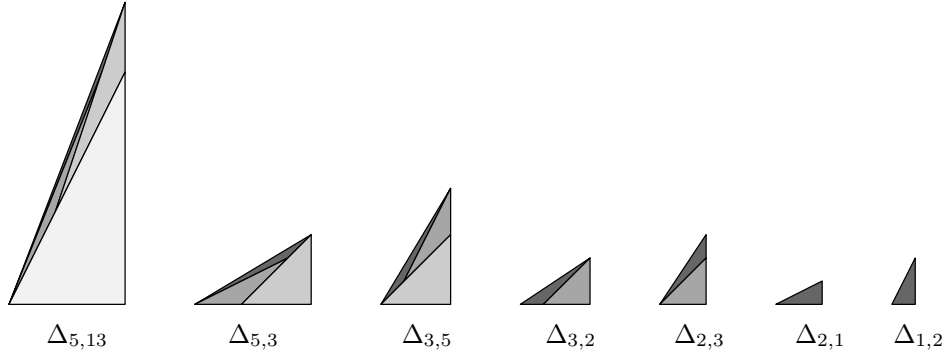


Figure 8: Similarly to our recursive description of $S_{5,13}$ (see Figure 5), we can apply Lemma 3.12 recursively to partition $\Delta_{5,13}$ into triangles with integral slope. The different shadings indicate which triangle the different regions correspond to.

The advantage of using the second expression for $T'_{a,ka}$ in 3.12.4 will become clear in Section 6 where we use it to obtain a short rational function representing the generating function of the set of lattice points inside $T'_{a,ka}$. Note that we obtain a recursion formula for T' by replacing every occurrence of T in 3.12.1 and 3.12.2 with T' and replacing 3.12.3 with $T'_{a,1} = \{(\frac{k}{0}) \mid 1 \leq k \leq a\}$.

In [KNA94] Kanamaru et al. use a recursive procedure as in Lemma 3.12 to give an algorithm to enumerate the set of lattice points *on* a line segment. They go on to give an algorithm that enumerates lattice points inside triangles using the transformation A , however in this case they do not apply recursion and do not mention the partition given in Lemma 3.12.1. This partition however is observed by Balza-Gomez et al. in [BGMM99]. But as they are interested in giving an algorithm for computing the convex hull of lattice points strictly below a line segment, they do not work with the full set of lattice points $T_{a,b}$. In both cases no explicit recursion formula such as Lemma 3.12 is given.

4 Characterizations of Sturmian Sequences

In this section we state several characterizations of Sturmian sequences of rational numbers, i.e. sequences of the form $\underline{B}_{a,b}$ (or equivalently $B_{a,b}$ with $0 < b \leq a$). We will first motivate each characterization in a separate paragraph without proofs and then summarize them in Theorem 4.1. The proof of the theorem occupies Section 5.

Recursive Structure. The most important characterization of Sturmian sequences for our purposes is a recursive one. It is based on the concept of reduction presented in Section 3 that relates $\underline{B}_{a,b}$ to $\underline{B}_{a,b \bmod a}$ in a way reminiscent of the Euclidean Algorithm.

We first present the idea informally. Let $0 < a < b$ and consider the sequence $B_{b,a}$ and the related staircase $S_{b,a}$. An interval of $B_{b,a}$ of the form $10 \dots 0$ of length k corresponds to a corner $c \in S_{b,a}$ and $k-1$ points in $S_{b,a}$ at the same height as c . We call a maximal interval of the form $10 \dots 0$ a block. A block of $B_{b,a}$ corresponds to a row of $S_{b,a}$. If the block has length k , the row contains k points. The *block sequence* $m(B_{b,a})$ of $B_{b,a}$ is the sequence of block lengths of $B_{b,a}$. By the above observation the block sequence of $B_{b,a}$ is the row sequence of $S_{b,a}$ and by Lemma 3.9.2 and Corollary 3.5 the column sequence of $\bowtie S_{b,a} = S_{a,b}$ *up to shift*. This means that $m(B_{b,a}) \equiv B_{a,b}$ is Beatty and hence $\underline{m}(B_{b,a}) \equiv B_{a,b \bmod a}$ is Sturmian. See Figure 9

for an example. It turns out that this gives a recursive characterization of Sturmian sequences: a sequence s is Sturmian iff $m(s)$ is balanced and $\underline{m}(s)$ is Sturmian.

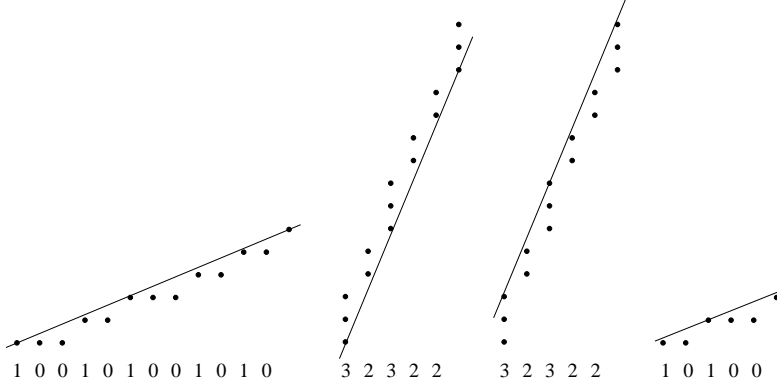


Figure 9: *Recursive structure of Sturmian sequences.* This figure shows several staircases $S_{a,b}$ and the corresponding parts of the associated Beatty sequence $B_{a,b}$. In each picture the first and last column correspond to the same element of $B_{a,b}$ modulo the minimal period. The first picture shows $S_{12,5}$ and $B_{12,5}$. $B_{12,5}$ records which columns contain corners, so the block sequence $m(B_{12,5})$ is just the row sequence of $S_{12,5}$. Reflecting at the main diagonal turns rows into columns, so $m(B_{12,5})$ is the column sequence of $\searrow S_{12,5} = S_{5,12}^-$, which is shown in the second picture. Note that in the first picture the part of the staircase that we show was chosen such that points from the first and the last column lie on the defining line - and this property is preserved under reflection at the main diagonal. Applying the reflection at the origin gets us to $\nwarrow \searrow S_{12,5} = S_{5,12}$, which is shown in the third picture along with its column sequence $B_{5,12}$. The reflection at the origin, however, reverses the column sequence. Fortunately Corollary 3.5 tells us that Sturmian sequences are invariant under reversals - up to a shift. This shift in the column sequence can be seen from the fact that in the third picture, the defining line does not pass through points of the first or last column anymore. From the first three pictures, we see that the column sequence of $S_{5,12}$ is just the row sequence of $S_{12,5}$ up to shift, i.e. $m(B_{12,5}) \equiv B_{5,12}$. Now we can apply reduction. We pass from $S_{5,12}$ to $S_{5,2}$, as shown in the last picture, which gives us $\underline{m}(B_{12,5}) \equiv B_{5,2}$. This is a geometric illustration of the combinatorial fact that if s is Sturmian, so is $\underline{m}(s)$.

To make this precise we give the following definitions. Suppose $s = (s_n)_{n \in \mathbb{Z}}$ is a periodic 0,1-sequence with $\mathbb{P}(s) = a$ and $\mathbb{1}(s) = b$. Without loss of generality we can assume that $s_0 = 1$. Let $i_1, \dots, i_b \in \{0, \dots, a-1\}$ be the indices of the 1s, i.e. let $i_1 < i_2 < \dots < i_b$ with $s_{i_j} = 1$ for all $1 \leq j \leq b$. Put $i_{b+1} := a$. Then the j -th block in (s_0, \dots, s_{a-1}) is $s|_{[i_j, i_{j+1}-1]}$ and the length of the j -th block is $m_j := i_{j+1} - i_j$. The *block sequence* $m(s)$ is the infinite periodic sequence generated by (m_1, \dots, m_b) . Note that this definition determines $m(s)$ only up to shift, which suffices for our purposes. On equivalence classes of sequences up to shift, m is an injective function, i.e. $s_1 \equiv s_2 \Leftrightarrow m(s_1) \equiv m(s_2)$. We call a sequence *block balanced* if it is balanced and its block sequence is balanced. In this case we can consider the reduced block sequence $\underline{m}(s)$ which is again a 0,1-sequence. A sequence s is *recursively balanced*

- if $\mathbb{1}(s) = 1$, or
- if s is block balanced and $\underline{m}(s)$ is recursively balanced.

The characterization now is this:

A periodic 0,1-sequence is Sturmian if and only if it is recursively balanced.

Even Distribution of 0s and 1s. Common sense suggests that, as the staircase approximates a line, the 0s and 1s of the Sturmian sequence should be distributed as evenly as possible. The actual number of 1s in every interval should be as close as possible to the expected number of 1s. This can be made precise in the following way. On an interval of length l , a line with slope $\frac{b}{a}$ increases by $\frac{b}{a}l$. So the expected number of 1s in an interval of length l of $\underline{B}_{a,b}$ is $\frac{b}{a}l$, if $b < a$, and $\frac{b \bmod a}{a}l$ in general. As $\frac{b \bmod a}{a}l$ is in general not an integer, the best that can be hoped for is that for every interval I of length l the number of 1s contained in l is either $\lfloor \frac{b \bmod a}{a}l \rfloor$ or $\lceil \frac{b \bmod a}{a}l \rceil$, and indeed this is a necessary and sufficient characterization of Sturmian sequences. Formally, we say that the 1s in a periodic 0,1-sequence s are *evenly distributed* if for every interval $s|_{[x_0, x_1]}$

$$(4.1) \quad \text{ones}(s|_{[x_0, x_1]}) \in \left\{ \left\lfloor \frac{\mathbb{1}(s)}{\mathbb{P}(s)} \text{length}(s|_{[x_0, x_1]}) \right\rfloor, \left\lceil \frac{\mathbb{1}(s)}{\mathbb{P}(s)} \text{length}(s|_{[x_0, x_1]}) \right\rceil \right\}.$$

Note that if $z \in \mathbb{Z}$ and $r \in \mathbb{R}$, then $z \in \{\lfloor r \rfloor, \lceil r \rceil\}$ if and only if $z - 1 < r < z + 1$, so the condition

$$(4.2) \quad \text{ones}(s|_{[x_0, x_1]}) - 1 < \frac{\mathbb{1}(s)}{\mathbb{P}(s)} \text{length}(s|_{[x_0, x_1]}) < \text{ones}(s|_{[x_0, x_1]}) + 1$$

is equivalent to (4.1). If an interval $s|_{[x_0, x_1]}$ violates the left-hand inequality, then we say it contains too many 1s and if it violates the right-hand inequality, we say it contains too few 1s. The characterization, then, is this.

A periodic 0,1-sequence s is Sturmian if and only if the 1s in s are evenly distributed.

This characterization appears in [GLL78] and was later improved in [Fra05].

Symmetry. A different way to phrase that the 0s and 1s are distributed evenly would be to state that Sturmian sequences are symmetric. If symmetric is taken to mean invariant under reversals (up to shift), then this is a true statement (Corollary 3.5) - but insufficient to characterize Sturmian sequences.

However, Lemma 3.2 suggests a different notion of symmetry. If we start with a flat staircase $S_{a,b}$ with $0 < b < a$ and move the defining line downwards by a small amount, the resulting staircase $S_{a,b,r}$ will be a translate of $S_{a,b}$. Hence their column sequences are identical up to shift. But using Lemma 3.1, we see that if $0 > r \geq -\frac{1}{a}$ the only columns that differ are col_{ka} for $k \in \mathbb{Z}$: the single point in these columns has been moved down by one. The columns col_{ka} do not contain a corner anymore, whereas the columns col_{ka+1} do. In the Sturmian sequence, this translates to taking the corresponding interval 1,0 and replacing it with the interval 0,1.

This observation gives rise to the notion of swap symmetry. A periodic 0,1-sequence s is swap symmetric if there is a pair $(s_i, s_{i+1}) = (1, 0)$ such that if we replace this pair and all periodic copies of it by $(0, 1)$, we obtain a sequence s' that is identical to s up to a shift. See Figure 10 for an example. Formally, given a periodic sequence s and $i \in \mathbb{Z}$, we define the sequence $\text{swap}(s, i) := (\text{swap}(s, i)_n)_n$ by

$$\text{swap}(s, i)_n = \begin{cases} s_n - 1 & \text{if } n \equiv i \pmod{\mathbb{P}(s)} \\ s_n + 1 & \text{if } n \equiv i + 1 \pmod{\mathbb{P}(s)} \\ s_n & \text{otherwise} \end{cases}.$$

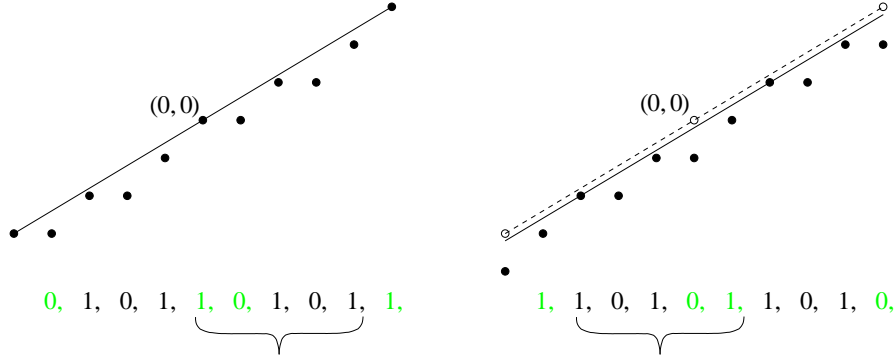


Figure 10: *Swap symmetry of Sturmian sequences.* Consider the staircase $S_{5,3}$ and the corresponding Sturmian sequence $B_{5,3}$, which records the columns of $S_{5,3}$ that contain a corner. If we move the line defining $S_{5,3}$ downwards by a small amount, then only those columns change in which there was a point on the line $L_{5,3}$. These columns do not contain corners any more - the corners move one column to the right. So in the sequence $B_{5,3}$ this corresponds to replacing an interval 10 with 01, i.e. by swapping a 1 and a 0. The digits that are swapped are highlighted in the figure. Now, by Lemma 3.2 translating the defining line only shifts the column sequence of $C_{5,3}$. So the sequence obtained from $B_{5,3}$ by swapping is again $B_{5,3}$ up to shift. This is shown by the braces, which indicate minimal periods of both sequences.

We call a periodic sequence s *swap symmetric* if there exists an $i \in \mathbb{Z}$ such that $s \equiv \text{swap}(s, i)$. Note that if $s \equiv \text{swap}(s, i)$ then

$$\begin{aligned} s_i &= \text{swap}(s, i)_{i+1} = s_{i+1} + 1 \\ s_{i+1} &= \text{swap}(s, i)_i = s_i - 1 \end{aligned}$$

as the number of entries $0 \leq k < \mathbb{P}(s)$ such that $s_k = c$ cannot change under a swap for any constant $c \in \mathbb{N}$, for swap symmetric s .

This property characterizes Sturmian sequences:

A periodic 0,1-sequence is Sturmian if and only if it is swap-symmetric.

Having motivated the three characterizations, we can now state the theorem.

Theorem 4.1. *Let $s = (s_n)_{n \in \mathbb{Z}}$ be a periodic 0,1-sequence with $\mathbb{P}(s) = a$ and $1(s) = b \geq 1$. Then the following are equivalent:*

- (i) $s \equiv \underline{B}_{a,b}$.
- (ii) s is recursively balanced.
- (iii) The 1s in s are evenly distributed.
- (iv) s is swap symmetric.

$(i) \Rightarrow (iii)$ is easy to prove (see next section) and $(iii) \Rightarrow (i)$ appears in [GLL78], although the concept of “nearly linear” sequences used in [GLL78] differs slightly from 4.1(iii). The connection between these definitions¹ is made in [Fra05], where the result from [GLL78] is extended. In both cases the focus lies on the more general

¹In [Fra05] sequences with an even distribution of 1s are called “balanced”.

case of lines with irrational slope. The proofs given in these two sources differ from the proofs we present in Section 5.

As far as we know the concepts of recursively balanced and swap symmetric sequences do not appear in the prior literature.

5 Proof of the Characterizations

We first show the recursive characterization, i.e. that a periodic 0,1-sequence is Sturmian if and only if it is recursively balanced. To that end we first prove two lemmas.

Lemma 5.1. *If $0 < a < b$, then $m(B_{b,a}) \equiv B_{a,b}$.*

Proof. Let $0 < a < b$. A maximal interval of $B_{b,a}$ of the form $1, 0, \dots, 0$ corresponds to a row of $S_{b,a}$, so

$$m(B_{b,a}) \equiv (|\text{row}_n(S_{b,a})|)_n = (|\text{col}_n(S_{a,b}^-)|)_n \equiv (|\text{col}_n(S_{a,b})|)_n = B_{a,b}$$

where we use Lemma 3.3 in the second and Lemma 3.2 in the third step. \square

From this we also get that for $a > b$ (i.e. flat staircases) the sequence $m(B_{a,b})$ is balanced, and thus Sturmian sequences are block balanced.

Lemma 5.2. *If s is a block balanced 0,1-sequence, then s is Sturmian if and only if $\underline{m}(s)$ is Sturmian.*

Proof. As s is a block balanced 0,1-sequence

$$\begin{aligned} s &\equiv B_{b,a} \text{ for some } 0 < a < b \\ \Leftrightarrow m(s) &\equiv B_{a,b} \text{ for some } 0 < a < b \\ \Leftrightarrow \underline{m}(s) &\equiv \underline{B}_{a,b} \text{ for some } 0 < a < b \end{aligned}$$

where the first equivalence holds by Lemma 5.1 and the fact that m is injective and the second equivalence holds by Lemma 3.6. \square

Proof of Theorem 4.1: (i) \Leftrightarrow (ii). The proof is by induction on $\mathbb{1}(s)$. If $\mathbb{1}(s) = 1$, the statement holds. For the induction step, we have the following equivalences:

$$\begin{aligned} &s \text{ is recursively balanced} \\ \Leftrightarrow &s \text{ is block balanced and} \\ &\underline{m}(s) \text{ is recursively balanced} \\ \Leftrightarrow &s \text{ is block balanced and} \\ &\underline{m}(s) \text{ is Sturmian} \\ \Leftrightarrow &s \text{ is block balanced and} \\ &s \text{ is Sturmian} \\ \Leftrightarrow &s \text{ is Sturmian} \end{aligned}$$

Here we use Lemma 5.2 in the third step. Note that the induction terminates in the case $\mathbb{1}(s) = 1$ since if $\mathbb{1}(s) > 1$, then s has blocks of different sizes and so $\mathbb{1}(s) > \mathbb{1}(\underline{m}(s)) \geq 1$. \square

Now we turn to the proof of the characterization that a periodic 0,1-sequence s is Sturmian if and only if the 1s in s are evenly distributed. One direction is easy to show.

Proof of Theorem 4.1: (i) \Rightarrow (iii). Let $B_{a,b}$ be a Sturmian sequence (i.e. $a > b$) and let $B_{a,b}|_{[x_0, x_1]}$ be any interval of $B_{a,b}$. Using (2.1) and the fact that

$$\frac{b}{a}(x_1 - x_0 + 1) = \frac{\mathbb{1}(B_{a,b})}{\mathbb{P}(B_{a,b})} \text{length}(B_{a,b}|_{[x_0, x_1]})$$

we obtain

$$(5.1) \quad \text{ones}(B_{a,b}|_{[x_0, x_1]}) = \frac{\mathbb{1}(B_{a,b})}{\mathbb{P}(B_{a,b})} \text{length}(B_{a,b}|_{[x_0, x_1]}) + \left\{ \frac{b}{a}(x_0 - 1) \right\} - \left\{ \frac{b}{a}x_1 \right\}.$$

So, since $0 \leq \left\{ \frac{b}{a}(x_0 - 1) \right\} < 1$ and $0 \leq \left\{ \frac{b}{a}x_1 \right\} < 1$ and both terms appear in (5.1) with opposite signs,

$$\text{ones}(B_{a,b}|_{[x_0, x_1]}) - 1 < \frac{\mathbb{1}(B_{a,b})}{\mathbb{P}(B_{a,b})} \text{length}(B_{a,b}|_{[x_0, x_1]}) < \text{ones}(B_{a,b}|_{[x_0, x_1]}) + 1.$$

□

To show the other direction, we make use of the first characterization. We show that if the 1s in s are evenly distributed, then s is recursively balanced.

Proof of Theorem 4.1: (iii) \Rightarrow (ii). Let s be a periodic 0,1-sequence in which the 1s are evenly distributed. We use induction on $\mathbb{1}(s)$. If $\mathbb{1}(s) = 1$, then by definition s is recursively balanced. For the induction step, we assume $\mathbb{1}(s) > 1$ and show that s is block balanced and the 1s in $\underline{m}(s)$ are evenly distributed. Then we can apply the induction hypothesis to obtain that $\underline{m}(s)$ and hence s is recursively balanced.

Step 1: s is block balanced. If there were blocks of zeros in s that differed in length by at least two, then we could find intervals u and v of the same length l , such that u contains two 1s and v contains none. But then $\{0, 2\} \subseteq \left\{ \left\lfloor \frac{\mathbb{1}(s)}{\mathbb{P}(s)}l \right\rfloor, \left\lceil \frac{\mathbb{1}(s)}{\mathbb{P}(s)}l \right\rceil \right\}$, which is impossible. So s is block balanced at some $k \in \mathbb{N}$, $\underline{m}(s)$ is well defined and the following identities hold.

$$\begin{aligned} \mathbb{P}(\underline{m}(s)) &= \mathbb{1}(s) \\ \mathbb{1}(\underline{m}(s)) &= \mathbb{P}(s) - k\mathbb{1}(s) \end{aligned}$$

Step 2: The 1s in $\underline{m}(s)$ are evenly distributed. Briefly, the idea is this: if m' is an interval of $\underline{m}(s) =: m$ that has too many 1s, looking at the corresponding interval in s we will find many large blocks (i.e. many 0s) and so we can construct an interval s'' of s that has too few 1s. This gives a contradiction to the assumption that the 1s in s are evenly distributed.

Let $m' = \underline{m}(s)|_{[x_0, x_1]}$ be an interval of $\underline{m}(s)$. We have to show that

$$(5.2) \quad \text{ones}(m') - 1 < \frac{\mathbb{1}(m)}{\mathbb{P}(m)} \text{length}(m') < \text{ones}(m') + 1.$$

Assume to the contrary that m' violates (5.2).

We first argue that without loss of generality

$$(5.3) \quad \text{ones}(m') - 1 \geq \frac{\mathbb{1}(m)}{\mathbb{P}(m)} \text{length}(m'),$$

i.e. that m' contains too many 1s. Suppose m' contains too few 1s, i.e. $\text{ones}(m') + 1 \leq \frac{\mathbb{1}(m)}{\mathbb{P}(m)} \text{length}(m')$. Then choose $x_2 > x_1$ such that the length of the interval $m|_{[x_0, x_2]}$ is a multiple $\alpha\mathbb{P}(m)$, $\alpha \in \mathbb{N}$ of the period length. Then $\text{ones}(m|_{[x_0, x_2]}) = \alpha\mathbb{1}(m)$ as m

is periodic. Now the interval $m|_{[x_1+1, x_2]}$ has too many 1s, i.e. $\text{ones}(m|_{[x_1+1, x_2]}) - 1 \geq \frac{\mathbb{1}(m)}{\mathbb{P}(m)} \text{length}(m|_{[x_1+1, x_2]})$, as witnessed by the following computation:

$$\begin{aligned}
\text{ones}(m|_{[x_1+1, x_2]}) &= \text{ones}(m|_{[x_0, x_2]}) - \text{ones}(m') = \alpha \mathbb{1}(m) - \text{ones}(m') \\
&\geq \alpha \mathbb{1}(m) - \frac{\mathbb{1}(m)}{\mathbb{P}(m)} \text{length}(m') + 1 \\
&= \frac{\mathbb{1}(m)}{\mathbb{P}(m)} (\alpha \mathbb{P}(m) - \text{length}(m')) + 1 \\
&= \frac{\mathbb{1}(m)}{\mathbb{P}(m)} \text{length}(m|_{[x_1+1, x_2]}) + 1.
\end{aligned}$$

Each element of m' corresponds to a block of 0s in s , where we take the block to include the preceding 1 but not the succeeding 1. Taking all the blocks in s together that correspond to elements of m' we obtain an interval $s' = s|_{[y_0, y_1]}$ of s . Let $s'' = s|_{[y_0+1, y_1]}$ denote the interval obtained from s' by removing the first 1. Then the following identities hold.

$$\begin{aligned}
\text{length}(m') &= \text{ones}(s') = \text{ones}(s'') + 1 \\
\text{ones}(m') &= \text{length}(s') - k \text{ones}(s') = \text{length}(s'') + 1 - k(\text{ones}(s'') + 1)
\end{aligned}$$

By substituting these and the identities obtained in Step 1 into (5.3) we obtain

$$\text{length}(s'') + 1 - k(\text{ones}(s'') + 1) \geq \frac{\mathbb{P}(s) - k\mathbb{1}(s)}{\mathbb{1}(s)} (\text{ones}(s'') + 1) + 1$$

which by canceling terms implies $\text{length}(s'') \geq \frac{\mathbb{P}(s)}{\mathbb{1}(s)} (\text{ones}(s'') + 1)$ and therefore $\text{ones}(s'') + 1 \leq \frac{\mathbb{1}(s)}{\mathbb{P}(s)} \text{length}(s'')$. This means that s'' is an interval in s with too few 1s, contradicting the assumption that the 1s in s are evenly distributed. \square

Finally, we turn to the characterization that a periodic 0,1-sequence is Sturmian if and only if it is swap symmetric. In Section 4 we have already tried to motivate that Sturmian sequences are swap symmetric, and the proof indeed proceeds as suggested by Figure 10.

Proof of Theorem 4.1: (i) \Rightarrow (iv). Let $0 < b < a$ and let $s = B_{a,b} = (|\text{col}_n(C_{a,b})|)_n$. We claim that

$$\text{swap}(s, 0) = (|\text{col}_n(C_{a,b,-\frac{1}{a}})|)_n \equiv (|\text{col}_n(C_{a,b})|)_n = s$$

which completes the proof. The equivalence in the second step holds by Lemma 3.1. All that is left to show is why the first equality holds.

To this end we argue as follows (see Figure 10): First we observe that shifting the line down by $-\frac{1}{a}$ only changes those columns $\text{col}_x(S_{a,b})$ with $x \bmod a = 0$. More precisely $\text{col}_x(S_{a,b,-\frac{1}{a}}) = \text{col}_x(S_{a,b}) - \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ if $0 = x \bmod a$ and $\text{col}_x(S_{a,b,-\frac{1}{a}}) = \text{col}_x(S_{a,b})$ otherwise. Now we observe that a point v in a flat staircase is a corner if and only if $v - e_1$ is not in the staircase. As we know which columns changed, and that $a \geq 2$, this allows us determine where the corners are after the shift. If $x \bmod a = 0$, then $|\text{col}_x(C_{a,b,-\frac{1}{a}})| = 0$. If $x \bmod a = 1$, then $|\text{col}_x(C_{a,b,-\frac{1}{a}})| = 1$. Otherwise $\text{col}_x(C_{a,b,-\frac{1}{a}}) = \text{col}_x(C_{a,b})$. \square

To show that swap symmetric sequences are Sturmian, we first prove two lemmas. Note that in Lemma 5.3 we do not claim that if s is swap symmetric, then $m(s)$ is balanced. However we can show that $m(s)$ is swap symmetric.² We then observe

²Our definition of swap symmetry was phrased such that it can be applied to arbitrary periodic sequences.

in Lemma 5.4 that swap symmetric sequences are necessarily balanced and hence $m(s)$ is balanced.

Lemma 5.3. *Let s be a periodic 0,1-sequence. If s is swap-symmetric and $\mathbb{1}(s) > 1$, then $m(s)$ is swap-symmetric.*

Proof. We know that $\mathbb{P}(m(s)) = \mathbb{1}(s) > 1$. Let $0 \leq i < \mathbb{P}(s)$ be such that $\text{swap}(s, i) \equiv s$. Then $s_i = 1$. Say the block preceding s_i is the j -th block of s . Swapping at i makes all the k -th blocks of s with $k \equiv j \pmod{\mathbb{1}(s)}$ larger by one and all the k -th blocks of s with $k \equiv j+1 \pmod{\mathbb{1}(s)}$ smaller by one while leaving all other blocks unmodified. As $\mathbb{P}(m(s)) > 1$, this means that $\text{swap}(m(\text{swap}(s, i)), j) \equiv m(s)$ but by assumption $\text{swap}(s, i) \equiv s$, so there exists an $l \in \mathbb{Z}$ such that

$$\text{swap}(m(s), l) \equiv \text{swap}(m(\text{swap}(s, i)), j) \equiv m(s)$$

which means that $m(s)$ is swap symmetric. \square

Lemma 5.4. *If s is a periodic swap-symmetric sequence, then s is balanced.*

Proof. Assume to the contrary that s contains at least three different entries. Let $0 \leq i < \mathbb{P}(s)$ be such that $\text{swap}(s, i) \equiv s$. Then $s_i \neq s_{i+1}$. Let $a = s_i$ and $b = s_{i+1}$. Let $c \in \mathbb{Z}$ be such that there exists a $j \in \mathbb{Z}$ with $s_j = c$ but $a \neq c \neq b$. We now define the parameter $d(s)$, which is the sum of the distances of any occurrence of a in $\text{period}(s)$ to the closest preceding occurrence of c in s , i.e.

$$d(s) := \sum_{0 \leq k < \mathbb{P}(s), s_k = a} k - \max \{l < k \mid s_l = c\}.$$

Note that if $s \equiv s'$, then $d(s) = d(s')$ as shifting a sequence to the left or right does not affect the distances between occurrences of values. Now the swap at i interchanges the a at position i and the b at position $i+1$, which increases the distance of this occurrence of a to the previous c by 1 and leaves all other distances of an occurrence of a to a previous c unaffected. Hence $d(\text{swap}(s, i)) = d(s) + 1$ and so $\text{swap}(s, i) \not\equiv s$, which is a contradiction. \square

After these two lemmas, the proof that swap symmetric sequences are Sturmian is easy. Again we proceed by showing that swap symmetric sequences are recursively balanced.

Proof of Theorem 4.1: (iv) \Rightarrow (ii). Let s be a periodic 0,1-sequence that is swap symmetric. There is an index i at which we can swap, so $\mathbb{1}(s) > 0$. If $\mathbb{1}(s) = 1$, s is recursively balanced by definition. So we can assume $\mathbb{1}(s) > 1$. By Lemma 5.3 it follows that $m(s)$ is swap symmetric. By Lemma 5.4 it follows that $m(s)$ is balanced. Taking both together we conclude that s is block balanced and that $\underline{m}(s)$ is well defined and swap symmetric. By induction we infer that $\underline{m}(s)$ is recursively balanced. But if s is block balanced and $\underline{m}(s)$ is recursively balanced, then s is recursively balanced. \square

6 Application: Short Representations

It is a celebrated result by Barvinok [Bar94] that there is a polynomial time algorithm for counting the number of lattice points inside a given rational polytope when the dimension of the polytope is fixed. Note that if the dimension is an input variable, the problem gets *NP*-hard [GJ79]. For more about the algorithm see [DL05], [DLHTY04] and the textbook [Bar08].

The crucial ingredient of Barvinok's proof was his result that the set of lattice points in a simplicial cone of any fixed dimension can be expressed using a short generating function. In this section we give a new proof of this result for the special case of 2-dimensional cones (Theorem 6.1). A generalization of our proof to higher dimensions is not immediate, we hope, however, that such a generalization can be found in the future.

We consider the Laurent polynomial ring $\mathbb{K}[x_1^\pm, \dots, x_d^\pm]$. For a vector $m = (m_1, \dots, m_d) \in \mathbb{Z}^d$ we write $x^m := x_1^{m_1} \dots x_d^{m_d}$. This gives a bijection between \mathbb{Z}^d and the set of monomials in $\mathbb{K}[x_1^\pm, \dots, x_d^\pm]$. We can thus represent the set of lattice points in a polyhedron P by the generating function $f_P(x) = \sum_{m \in \mathbb{Z}^d \cap P} x^m$. If $P \cap \mathbb{Z}^d$ is large, this representation of f_P contains many terms. Using rational functions it is possible to find shorter representations of f_P . For example the generating function of all non-negative integral multiples of a vector m can be written as $\frac{1}{1-x^m}$, which allows us to express point sets like $\{0, m, \dots, km\}$ as $\frac{1-x^{(k+1)m}}{1-x^m}$. Developing these notions in detail is beyond the scope of this article. As references we recommend [BR07] and [Bar08]. However, we would like to point out, informally, how the algebraic operations on generating functions correspond to geometric operations: The sum of generating functions corresponds to the union of the respective sets. The product of generating functions corresponds to the Minkowski sum of the respective sets. Taking the product of a generating function and a monomial x^m thus corresponds to translation by m . Evaluating the generating function $f_P(x)$ at the values x^{m_1}, \dots, x^{m_d} for $m_1, \dots, m_d \in \mathbb{Z}^d$ corresponds to applying the linear map given by the matrix $A = (m_1 \dots m_d)$ that has the m_i as columns to the set P . As we already mentioned, for every 2-dimensional rational cone K in \mathbb{R}^2 there exists a lattice transform A such that $AK = \text{cone}((\frac{1}{0}), (\frac{a}{b}))$ for $a, b \in \mathbb{N}$ with $\gcd(a, b) = 1$. Barvinok showed a general version of the following theorem for cones of any dimension. We are going to give a new proof of this version for 2-dimensional cones (recall the definitions from the end of Section 3).

Theorem 6.1. *Let $a, b \in \mathbb{N}$ with $\gcd(a, b) = 1$. Let $K = \text{cone}((\frac{1}{0}), (\frac{a}{b}))$. Then f_K admits a representation as a rational function with $\mathcal{O}(\log a)$ terms and this representation can be computed in time polynomial in $\log a + \log b$.*

Proof. Step 1. We express f_K in terms of $f_{\Delta'_{a,b}}$. To this end we first note that

$$\text{cone}((\frac{1}{0}), (\frac{a}{b})) = \bigcup_{k \geq 0} k(\frac{a}{b}) + (\text{cone}((\frac{1}{0}), (\frac{a}{b})) \cap \mathbb{R} \times [0, b))$$

and

$$\begin{aligned} \mathbb{Z}^2 \cap \text{cone}((\frac{1}{0}), (\frac{a}{b})) \cap \mathbb{R} \times [0, b) &= \{(\frac{0}{0})\} \cup T'_{a,b} \\ &\cup \left(\left\{ (\frac{i}{0}) \mid i \geq a+1 \right\} + \left\{ (\frac{0}{j}) \mid 0 \leq j \leq b-1 \right\} \right). \end{aligned}$$

See Figure 11. In terms of generating functions this translates into

$$f_K(x) = \frac{1}{1-x_1^a x_2^b} \left(1 + f_{\Delta'_{a,b}}(x) + \frac{x_1^{a+1}}{1-x_1} \frac{1-x_2^b}{1-x_2} \right).$$

Here we express f_K using $f_{\Delta'_{a,b}}$ and a constant number of other terms. So it suffices to give a short expression of $f_{\Delta'_{a,b}}$.

Step 2. We use the recursion from Lemma 3.12 to give a short expression for $f_{\Delta'_{a,b}}$. Let $(c_n)_n$ be the sequence defined by $c_1 = b$, $c_2 = a$ and $c_{i+2} = c_i \bmod c_{i+1}$ and let j be the index such that $c_{j+1} = 1$ and $c_{j+2} = 0$. We express $f_{T'_{c_{j+1}, c_j}}$ in terms

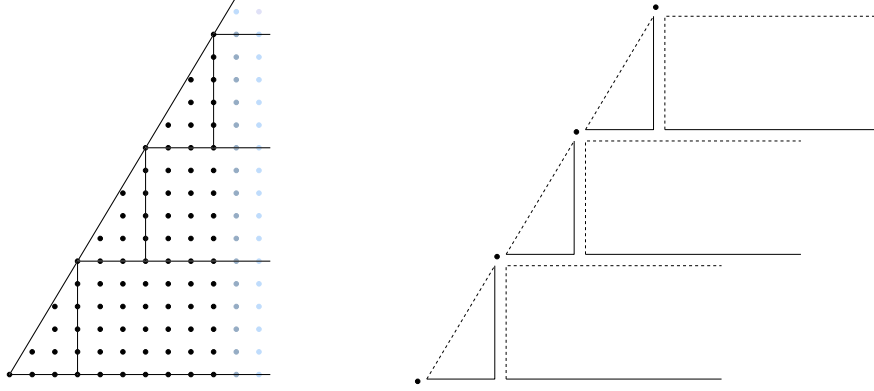


Figure 11: Expressing the lattice points in a cone in terms of triangles. The right picture shows the occurring shapes (dashed lines indicate open faces).

of $f_{T'_{c_{i+2}, c_{i+1}}}$, by applying first 3.12.1 and then 3.12.2 and 3.12.4.

$$\begin{aligned}
f_{T'_{c_{i+1}, c_i}}(x_1, x_2) &= f_{T'_{c_{i+1}, c_{i+2}}}(x_1 x_2^{c_i \operatorname{div} c_{i+1}}, x_2) \\
&\quad + f_{T'_{c_{i+1}, (c_i \operatorname{div} c_{i+1})c_{i+1}}}(x_1, x_2) \\
&= x_1^{c_{i+1}} x_2^{c_{i+2}} \cdot f_{T'_{c_{i+2}, c_{i+1}}}(x_2^{-1}, x_1^{-1} x_2^{-(c_i \operatorname{div} c_{i+1})}) \\
&\quad + \frac{1}{1-x_2} \cdot \left(\frac{1-x_1^{c_{i+1}+1}}{1-x_1} - \frac{1-x_1^{c_{i+1}+1} x_2^{(c_i \operatorname{div} c_{i+1})(c_{i+1}+1)}}{1-x_1 x_2^{c_i \operatorname{div} c_{i+1}}} \right).
\end{aligned}$$

We have thus expressed $f_{T'_{c_{i+1}, c_i}}$ using a constant number of other terms. We proceed in this fashion until we reach the case $f_{T'_{c_{j+1}, c_j}} = f_{T'_{1, c_j}}$ which we can solve directly using 3.12.3:

$$f_{T'_{1, c_j}} = x_1 \frac{1-x_2^{c_j}}{1-x_2}.$$

Step 3. The expression is short and can be computed in polynomial time as the Euclidean Algorithm is fast. By Lemma 3.10 the number of iterations required in step 2 is $\mathcal{O}(\log a)$. In each step we pick up a constant number of terms. So the total number of terms in the final expression is $\mathcal{O}(\log a)$. The algorithm runs in time polynomial in $\log a + \log b$ as the numbers $c_{i+2} = c_i \bmod c_{i+1}$ and $c_i \operatorname{div} c_{i+1}$ can be computed in time polynomial in $\log c_i + \log c_{i+1}$. \square

This proof of Theorem 6.1 differs from Barvinok's. Barvinok gives a *signed* decomposition of a cone into *unimodular* cones. We give a *positive* decomposition of the triangle $T'_{a,b}$ into triangles $T'_{c_{i+1}, (c_i \operatorname{div} c_{i+1})c_{i+1}}$ that are not unimodular but easy to describe, i.e. using a constant number of terms.

In this context “positive” means that the 2-dimensional triangle $T'_{a,b}$ is written as a disjoint union of half-open 2-dimensional triangles $T'_{a,b}$. This does not mean that the numerator of the rational function has only positive coefficients. Negative coefficients appear in the “easy” description of the triangles $T'_{c_{i+1}, (c_i \operatorname{div} c_{i+1})c_{i+1}}$.

Lemma 3.11 can be used to obtain a short representation of the generating function of the lattice points in the fundamental parallelepiped of any rational cone in the plane. We implement this idea in the proof of Theorem 7.1 in Section 7. This representation can also be used to give an alternative proof of Theorem 6.1. Again the representation is positive in the sense that the set of lattice points in

the fundamental parallelepiped is expressed as a disjoint union of Minkowski sums of intervals. But of course still negative coefficients appear as they appear in the representation of intervals. As opposed to the representation based on triangles, the representation based on fundamental parallelepipeds relies on taking products; so with this approach expanding the products in the numerators leads to an expression that is not short any more.

It is also possible to give a recursion similar to 3.9, 3.11 and 3.12 directly for cones. However, in this case the recursion does require us to take differences of sets and we do not obtain a “positive” decomposition. Nonetheless the recursion differs from the one based on the continued fraction expansion of $\frac{b}{a}$ given in [Bar08, Chapter 15].

7 Application: Dedekind-Carlitz Polynomials

Given $0 < a, b \in \mathbb{N}$ with $\gcd(a, b) = 1$, Carlitz introduced the following polynomial generalization of Dedekind sums, which Beck, Haase and Matthews in [BHM08] call the *Dedekind-Carlitz polynomial*:

$$c_{a,b}(x, y) := \sum_{k=1}^{a-1} x^{k-1} y^{\lfloor \frac{b}{a} k \rfloor}.$$

For a brief overview of the history of and literature about Dedekind sums and the Dedekind-Carlitz polynomial, we refer to [BHM08]. There also the relationship between Dedekind-Carlitz polynomials and the fundamental parallelepipeds of cones (see below) is established. Appealing to Barvinok’s Theorem, Beck, Haase and Matthews conclude that the Dedekind-Carlitz polynomial can be computed in polynomial time and must have a short representation,³ however they do not give such a short representation explicitly. Also they remark that Dedekind sums can be computed efficiently in the style of the Euclidean Algorithm and ask if such a recursive procedure also exists for Dedekind-Carlitz polynomials. In this section we use the recursion for the lattice points inside a fundamental parallelepiped developed in Section 3 to give an explicit recursion formula that allows one to compute short representations of Dedekind-Carlitz polynomials in the style of the Euclidean Algorithm.

We first observe that $c_{a,b}$ is the generating function of the set

$$\left\{ z \in \mathbb{Z}^2 \mid z_1 = \left\lfloor \frac{b}{a} z_2 \right\rfloor, 1 \leq z_1 \leq a-1 \right\} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \Pi_{\downarrow, a, b}^\circ - \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

which is just a translate of the set of lattice points in the open fundamental parallelepiped $\Pi_{\downarrow, a, b}^\circ$. Hence the recursion given in Lemma 3.11 can be used to give a recursion formula for Dedekind-Carlitz sums in the spirit of the Euclidean Algorithm. To this end, we use $d_{a,b}^\downarrow(x, y)$ and $d_{a,b}^\rightarrow(x, y)$ to denote the generating functions of the sets $\Pi_{\downarrow, a, b}^\circ$ and $\Pi_{\rightarrow, a, b}^\circ$, respectively. So

$$\begin{aligned} d_{a,b}^\downarrow(x, y) &= \sum_{k=1}^{a-1} x^k y^{\lfloor \frac{b}{a} k \rfloor}, \\ d_{a,b}^\rightarrow(x, y) &= \sum_{k=1}^{b-1} x^{\lceil \frac{a}{b} k \rceil} y^k. \end{aligned}$$

Now, by simply translating the geometric operations into the language of generating functions, we obtain the following theorem. In [BHM08] this result is derived from Barvinok’s Theorem. We give an explicit recursion formula in the proof.

³In [BHM08] this is argued even for higher-dimensional Dedekind-Carlitz polynomials.

Theorem 7.1. *Let $0 < a, b \in \mathbb{N}$ with $\gcd(a, b) = 1$. Then $c_{a,b}$ admits a representation as a rational function with $\mathcal{O}(\log a)$ terms and this representation can be computed in time polynomial in $\log a + \log b$.*

As was said before, the representation we obtain is “positive” in the sense that we build a partition of the set $\Pi_{\downarrow, a, b}^\circ$ using Minkowski sums and disjoint unions of intervals. It is not positive in the sense that all coefficients appearing the representation are positive, as the representations of intervals that we use contain coefficients with opposite signs.

It is important to stress that the representation we obtain makes heavy use of Minkowski sums of intervals. In the language of generating functions, this corresponds to taking products of expressions of the form $\frac{1-x^{ku}}{1-x^u}$ for $k \in \mathbb{N}$ and $u \in \mathbb{Z}^2$. Expanding the numerators of these products by applying the distributive law may lead to a numerator with a number of summands exponential in the number of factors of the product. So the expression we obtain is only short, if products are not expanded. We note that this problem does not occur with the representation we used in the proof of Theorem 6.1.

Proof. First we note that $c_{a,b}(x, y) = x^{-1}d_{a,b}^\downarrow(x, y)$. Now we construct a short representation of $d_{a,b}^\downarrow(x, y)$ by applying Lemma 3.11 inductively. To that end let $(c_n)_n$ be the sequence defined by $c_1 = b$, $c_2 = a$ and $c_{i+2} = c_i \bmod c_{i+1}$ and let j be the index such that $c_{j+1} = 1$ and $c_{j+2} = 0$. Such a j exists because $\gcd(a, b) = 1$. By Lemma 3.11.2 we can assume without loss of generality $c_1 > c_2$. Then for all $i \geq 1$

$$\begin{aligned} d_{c_{i+1}, c_i}^\downarrow(x, y) &= d_{c_{i+1}, c_{i+2}}^\downarrow(xy^{c_i \operatorname{div} c_{i+1}}, y) \\ d_{c_{i+1}, c_i}^\rightarrow(x, y) &= \frac{1 - y^{-(c_i \operatorname{div} c_{i+1})}}{1 - y^{-1}} d_{c_{i+1}, c_{i+2}}^\downarrow(xy^{c_i \operatorname{div} c_{i+1}}, y) \\ &\quad + y^{-(c_i \operatorname{div} c_{i+1})} d_{c_{i+1}, c_{i+2}}^\rightarrow(xy^{c_i \operatorname{div} c_{i+1}}, y) \\ &\quad + x^a y^b \frac{y^{-1} - y^{-(c_i \operatorname{div} c_{i+1})}}{1 - y^{-1}} \end{aligned}$$

and

$$d_{c_i, c_{i+1}}^\rightarrow(x, y) = x^a y^b d_{c_{i+1}, c_i}^\downarrow(-y, -x) \quad \text{and} \quad d_{c_i, c_{i+1}}^\downarrow(x, y) = x^a y^b d_{c_{i+1}, c_i}^\rightarrow(-y, -x).$$

Together with

$$d_{c_j, c_{j+1}}^\rightarrow(x, y) = 0 \quad \text{and} \quad d_{c_j, c_{j+1}}^\downarrow(x, y) = \frac{x - x^a}{1 - x}$$

this gives us a recursion formula for $d_{a,b}^\downarrow(x, y)$. In each step of the recursion we pick up only a constant number of terms and by Lemma 3.10 we need only $\mathcal{O}(\log a)$ steps, so the resulting representation has only $\mathcal{O}(\log a)$ terms. As standard arithmetic operations can be computed in time polynomial in the input length, the algorithm runs in time polynomial in $\log a + \log b$. \square

Note that by using products, one can give a representation of the lattice points in an interval of length n with $\mathcal{O}(\log n)$ many terms and *without using rational functions*. Using such a representation, the above proof gives a representation with $\mathcal{O}(\log^2 a)$ terms in time polynomial in $\log a + \log b$. Moreover this representation then is positive in that every coefficient appearing in this expression has a positive sign.

8 Application: Theorem of White

To conclude this paper, we give a partly new proof for a theorem of White [Whi64, pp.390-394], characterizing lattices tetrahedra containing no lattice points but the vertices. Several proofs appeared over the years, e.g. by Noordzij [Noo81], Scarf [Sca85] (based partly on work by R. Howe) and recently Reznick [Rez06], who also gives an overview of the history of this theorem. Furthermore his proof has the advantage that it keeps track of the vertices, at the cost of geometric transparency. We construct our proof based on ideas in [Sca85] and mainly [Rez06].

For $(a, b, n) \in \mathbb{Z}^3$ we define the tetrahedron $T_{a,b,n}$ as

$$\text{conv} \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} a \\ b \\ n \end{pmatrix} \right\}.$$

This should not be confused with the point set $T_{a,b}$ defined in Section 3. As we will not use the latter any more, no ambiguities should arise.

A “hidden” parameter, as Reznick writes, is $c = n - a - b + 1$ (although he considers a slightly different c). We will see that c plays a role equal to the ones of a and b in $T_{a,b,n}$. Also note that $a + b + c = n + 1$.

We call two tetrahedra T and T' *equivalent* ($T \approx T'$), if there is an affine lattice transformation which takes the vertices of T to the vertices of T' .

A lattice simplex T is *clean* if there are no non-vertex lattice points on the boundary. If there are also no lattice points in the interior of T (i.e. the vertices are the only lattice points), then we call T *empty*.

Theorem 8.1 (White). *A lattice tetrahedron T is empty if and only if it is equivalent to $T_{0,0,1}$ or to some $T_{1,d,n}$, where $\gcd(d, n) = 1$ and $1 \leq d \leq n - 1$.*

Proof of Theorem 8.1 (Necessity). As we easily see, $T_{0,0,1}$ is empty. So we consider $T_{1,d,n}$, where $\gcd(d, n) = 1$ and $1 \leq d \leq n - 1$. Let $w \in \mathbb{Z}^3 \cap T_{1,d,n}$. As the first coordinate of all vertices of $T_{1,d,n}$ is either 0 or 1, we know $w_1 \in \{0, 1\}$.

If $w_1 = 0$, then $w \in \text{conv} \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ and w is a vertex.

If $w_1 = 1$, then $w \in \text{conv} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ d \\ n \end{pmatrix} \right\}$. As $\gcd(d, n) = 1$ the vector $\begin{pmatrix} 0 \\ d \\ n \end{pmatrix}$ is primitive and so w is again a vertex.

Therefore $T_{1,d,n}$ and any equivalent tetrahedron is empty. This proves that lattice tetrahedra of the form $T_{0,0,1}$ or $T_{1,d,n}$ with $\gcd(d, n) = 1$ and $1 \leq d \leq n - 1$ are necessarily empty. \square

To show the sufficiency, we use the following theorem by Reeve. A nice proof can be found in [Rez06, pp.5-6].

Theorem 8.2 ([Ree57]). *The lattice tetrahedron T is clean if and only if $T \approx T_{0,0,1}$ or $T \approx T_{a,b,n}$, where*

$$n \geq 2, \quad 0 \leq a, b \leq n - 1 \quad \text{and} \quad \gcd(a, n) = \gcd(b, n) = \gcd(1 - a - b, n) = 1.$$

We will now prove the sufficiency to motivate the rest of the section, where we anticipate the results that are stated and proved only afterwards.

Proof of Theorem 8.1 (Sufficiency). If T is an empty lattice tetrahedron, then in particular it is clean and thus by Theorem 8.2 equivalent to $T_{0,0,1}$ or some $T_{a,b,n}$. If

$T \approx T_{0,0,1}$, we are done. Otherwise $T_{a,b,n} \approx T$ is of course also empty and therefore fulfills the conditions for Lemma 8.4.

This in turn enables us to use Lemma 8.5 which tells us that one of a, b, c equals 1. Finally, we can see by Lemma 8.3 that we can choose the order of the coordinates freely, so we get $T \approx T_{1,d,n}$, where $\gcd(d, n) = 1$ and $1 \leq d \leq n-1$, again by Theorem 8.2. \square

It turns out to be useful to describe a clean tetrahedron $T_{a,b,n}$ in a slightly different way:

Lemma 8.3. *Let $T_{a,b,n}$ be empty and $0 \leq a, b \leq n-1$. Then*

$$T_{a,b,n} \approx \text{conv} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} a \\ b \\ c \end{pmatrix} \right\}, \quad \text{and } c \geq 1.$$

In the next two proofs we follow mostly [Sca85, pp.411f].

Proof. If $n < a + b$, then

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \alpha_1 \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \alpha_4 \begin{pmatrix} a \\ b \\ n \end{pmatrix}, \text{ where}$$

$$\alpha_4 = \frac{1}{n}, \quad \alpha_3 = 1 - \frac{b}{n}, \quad \alpha_2 = 1 - \frac{a}{n}, \text{ and } \alpha_1 = 1 - \alpha_2 - \alpha_3 - \alpha_4 = \frac{a+b-n-1}{n}.$$

But this means that $0 \leq \alpha_1 < 1$ and $0 < \alpha_2, \alpha_3, \alpha_4 < 1$, and thus $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \in T_{a,b,n}$, contradicting the assumption that $T_{a,b,n}$ is empty.

So we know that $n \geq a + b$, which proves $c \geq 1$. The affine lattice transformation

$$x \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} x + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

gives us the desired form for $T_{a,b,n}$. \square

The key ingredient for the rest of the proof of Theorem 8.1 is to look at the Beatty sequences for $\frac{a}{n}$, $\frac{b}{n}$ and $\frac{c}{n}$ simultaneously. For this purpose let us define the sum of the sequences, i.e.

$$\begin{aligned} f(k) &:= B_{n,a}(k) + B_{n,b}(k) + B_{n,c}(k) \\ &= \left\lfloor \frac{a}{n}k \right\rfloor + \left\lfloor \frac{b}{n}k \right\rfloor + \left\lfloor \frac{c}{n}k \right\rfloor - \left\lfloor \frac{a}{n}(k-1) \right\rfloor - \left\lfloor \frac{b}{n}(k-1) \right\rfloor - \left\lfloor \frac{c}{n}(k-1) \right\rfloor. \end{aligned}$$

This function has a strong connection to Theorem 8.1 as we see next.

Lemma 8.4. *If $T_{a,b,n}$ is empty, then $f(k) = 1$ for $k = 2, \dots, n-1$.*

Proof. Let us first define the function $g(k) := \left\lceil \frac{a}{n}k \right\rceil + \left\lceil \frac{b}{n}k \right\rceil + \left\lceil \frac{c}{n}k \right\rceil$. An easy computation verifies that for a, b, c relatively prime to n and $2 \leq k \leq n-1$, we have $f(k) = g(k) - 3 - (g(k-1) - 3) = g(k) - g(k-1)$.

We will now show that $g(k) = k + 2$ for $k = 1, \dots, n-1$ and also that a, b, c are relatively prime to n .

Suppose that $g(k) \leq k + 1$ for some k . Then we can define a lattice point $\begin{pmatrix} p \\ q \\ r \end{pmatrix}$ with

$$p \geq \left\lceil \frac{a}{n}k \right\rceil, \quad q \geq \left\lceil \frac{b}{n}k \right\rceil, \quad r \geq \left\lceil \frac{c}{n}k \right\rceil,$$

and $p + q + r = k + 1$. But then we find $\alpha_1, \dots, \alpha_4$ with

$$\begin{pmatrix} p \\ q \\ r \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_4 \end{pmatrix}, \quad \text{where}$$

$$\alpha_1 = p - \frac{a}{n}k, \quad \alpha_2 = q - \frac{b}{n}k, \quad \alpha_3 = r - \frac{c}{n}k, \quad \alpha_4 = \frac{1}{n}k \quad \text{and}$$

$$\begin{aligned} \sum_{i=1}^4 \alpha_i &= p + q + r - \frac{a}{n}k - \frac{b}{n}k - \frac{c}{n}k + \frac{k}{n} \\ &= k + 1 - \frac{(a + b + c - 1)k}{n} = k + 1 - k \\ &= 1. \end{aligned}$$

As this means that $\begin{pmatrix} p \\ q \\ r \end{pmatrix}$ is in $\text{conv} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} a \\ b \\ c \end{pmatrix} \right\} \approx T_{a,b,n}$, we have a contradiction.

So we know that $g(k) \geq k + 2$.

If one of a, b, c is not relatively prime to n , say it is c , then $\frac{c}{n}k \in \mathbb{Z}$ for some $k \in \{1, \dots, n-1\}$. So we get

$$\begin{aligned} g(k) + g(n-k) &= \left\lceil \frac{a}{n}k \right\rceil + \left\lceil \frac{b}{n}k \right\rceil + \left\lceil \frac{c}{n}k \right\rceil + \left\lceil a - \frac{a}{n}k \right\rceil + \left\lceil b - \frac{b}{n}k \right\rceil + \left\lceil c - \frac{c}{n}k \right\rceil \\ &= \left\lceil \frac{a}{n}k \right\rceil + \left\lceil \frac{b}{n}k \right\rceil + \left\lceil \frac{c}{n}k \right\rceil + a - \left\lfloor \frac{a}{n}k \right\rfloor + b - \left\lfloor \frac{b}{n}k \right\rfloor + c - \left\lfloor \frac{c}{n}k \right\rfloor, \end{aligned}$$

and by the assumption on c this is at most $a + b + c + 2 = n + 3$. But then either $g(k) \leq k + 1$ or $g(n-k) \leq n - k + 1$, which cannot be true for an empty $T_{a,b,n}$.

If they are all relatively prime to n , then $g(k) + g(n-k) = a + b + c + 3 = n + 4$. Together with $g(k) \geq k + 2$ and $g(n-k) \geq n - k + 2$ we get the desired equality.

We have now seen that $g(k) = k + 2$ for $k = 1, \dots, n-1$. Together with $f(k) = g(k) - g(k-1)$ this implies $f(k) = 1$ for $k = 2, \dots, n-1$. \square

This is all we need to finish the proof:

Lemma 8.5. *If $f(k) = 1$ for $k = 2, \dots, n-1$, then at least one of a, b or c equals 1.*

This is the new part of the proof. It builds onto the observations about Sturmian sequences developed in the first half of this article. In particular we make use of the fact that Sturmian sequences are block balanced, and, more generally, that the 1s in a Sturmian sequence are evenly distributed. See Theorem 4.1(iii).

Proof. Suppose not. Without loss of generality $c < a, b$. We consider the intervals $B_{n,a}|_{[1,n]}$, $B_{n,b}|_{[1,n]}$ and $B_{n,c}|_{[1,n]}$. As $c \geq 2$, there is another 1 in $B_{n,c}|_{[1,n]}$ apart from $B_{n,c}(n) = 1$. Let m be the position of the 1 preceding $B_{n,c}(n) = 1$, i.e. $1 < m < n$ such that $B_{n,c}(m) = 1$ and $B_{n,c}(k) = 0$ for $m < k < n$. Because $f(k) = 1$ for all $2 \leq k \leq n-1$, we know that $B_{n,a}(m) = B_{n,b}(m) = 0$.

At this point we make a table listing the values of the three intervals at the positions $1, \dots, n$, filling in the values that we know and marking the values that we have not yet determined with $*$.

$$\begin{array}{cccccccccc}
1 & 2 & \dots & m-1 & m & m+1 & \dots & n-1 & n \\
B_{n,a}|_{[1,n]} = & 0 & * & \dots & * & 0 & * & \dots & * & 1 \\
B_{n,b}|_{[1,n]} = & 0 & * & \dots & * & 0 & * & \dots & * & 1 \\
B_{n,c}|_{[1,n]} = & 0 & * & \dots & * & 1 & 0 & \dots & 0 & 1
\end{array}$$

Because $f(n-1) = 1$ we know that either $\underline{B}_{n,a}(n-1) = 1$ or $\underline{B}_{n,b}(n-1) = 1$ and we may assume it is $\underline{B}_{n,a}(n-1) = 1$. We are now going to apply the following argument over and over again. By Theorem 4.1(iii) we know that if we find an interval of length l in a Sturmian sequence that contains t 1s, then any other interval of length l in the same sequence has to contain at least $t-1$ and at most $t+1$ 1s. In this case $\text{ones}(B_{n,a}|_{[n-1,n]}) = 2$ and so both $\text{ones}(B_{n,a}|_{[m-1,m]}) \geq 1$ and $\text{ones}(B_{n,a}|_{[m,m+1]}) \geq 1$, which means $B_{n,a}(m-1) = B_{n,a}(m+1) = 1$ and consequently $B_{n,b}(m-1) = B_{n,b}(m+1) = 0$. Now our table looks as follows.

$$\begin{array}{cccccccccc}
1 & 2 & \dots & m-1 & m & m+1 & \dots & n-1 & n \\
B_{n,a}|_{[1,n]} = & 0 & * & \dots * & 1 & 0 & 1 & * \dots * & 1 & 1 \\
B_{n,b}|_{[1,n]} = & 0 & * & \dots * & 0 & 0 & 0 & * \dots * & 0 & 1 \\
B_{n,c}|_{[1,n]} = & 0 & * & \dots * & 0 & 1 & 0 & \dots & 0 & 1
\end{array}$$

But now $\text{ones}(B_{n,b}|_{[m-1,m+1]}) = 0$ and so $B_{n,b}(n-2) = 0$ and $B_{n,a}(n-2) = 1$. This gives $\text{ones}(B_{n,a}|_{[n-2,n]}) = 3$ which allows us to deduce $B_{n,a}(m+2) = 1$ and $B_{n,b}(m+2) = 0$. Then we have $\text{ones}(B_{n,b}|_{[m-1,m+2]}) = 0$ and so $B_{n,b}(n-3) = 0$ and $B_{n,a}(n-3) = 1$, which gives $\text{ones}(B_{n,a}|_{[n-3,n]}) = 4$ and so $B_{n,a}(m+3) = 1$ and $B_{n,b}(m+3) = 0$. We continue this argument inductively until we have shown that $B_{n,b}(k) = 0$ for $m+1 \leq k \leq n-1$.

At this point both $B_{n,b}|_{[m-1,n-1]}$ and $B_{n,c}|_{[m+1,n-1]}$ are intervals of consecutive 0s, of length $n-m+1$ and $n-m-1$, respectively, where the latter is maximal. So the blocks of $B_{n,b}$ are strictly larger than the blocks of $B_{n,c}$. As the 1s in Sturmian sequences are evenly distributed, this implies $c = \mathbb{1}(B_{n,c}) > \mathbb{1}(B_{n,b}) = b$, in contradiction to our assumption. \square

Acknowledgements. We would like to thank Christian Haase for drawing our attention to White's Theorem and asking if our observations about Sturmian sequences can be used to find a simpler proof. We thank Victor Alvarez for pointing us to the work of Balza-Gomez et al. We also thank Matthias Beck for helpful discussions about our results and getting us interested in lattice points inside triangles in the first place.

References

- [Bar94] Alexander Barvinok, *A Polynomial Time Algorithm for Counting Integral Points in Polyhedra When the Dimension Is Fixed*, Mathematics of Operations Research **19** (1994), no. 4, 769–779.
- [Bar08] ———, *Integer Points in Polyhedra*, Zurich Lectures in Advanced Mathematics, European Mathematical Society, 2008.
- [BGMM99] Hélymar Balza-Gomez, Jean-Michel Moreau, and Dominique Michelucci, *Convex Hull of Grid Points below a Line or a Convex Curve*, Proceedings of the 8th International Conference on Discrete

- Geometry for Computer Imagery (London), Springer-Verlag, 1999, pp. 361–374.
- [BHM08] Matthias Beck, Christian Haase, and Asia R. Matthews, *Dedekind-carlitz polynomials as lattice-point enumerators in rational polyhedra*, Mathematische Annalen **341** (2008), no. 4, 945–961.
 - [BR07] Matthias Beck and Sinai Robins, *Computing the Continuous Discretely*, Springer-Verlag, 2007.
 - [Bro93] Tom C. Brown, *Descriptions of the Characteristic Sequence of an Irrational*, Canadian Mathematical Bulletin **36** (1993), no. 1, 15–21.
 - [DL05] Jesús A. De Loera, *The many aspects of counting lattice points in polytopes*, Mathematische Semesterberichte **52** (2005), no. 2, 175–195.
 - [DLHTY04] Jesús A. De Loera, Raymond Hemmecke, Jeremiah Tauzer, and Ruriko Yoshida, *Effective lattice point counting in rational convex polytopes*, Journal of Symbolic Computation **38** (2004), no. 4, 1273–1302.
 - [FMT78] Aviezri S. Fraenkel, Mordechai Mushkin, and Uzi Tassa, *Determination of $[n\theta]$ by its sequences of differences*, Canadian Mathematical Bulletin **21** (1978), 441–446.
 - [Fra05] Aviezri S. Fraenkel, *Complementary balanced words over an arbitrary finite alphabet*, <http://www.wisdom.weizmann.ac.il/~fraenkel/> (2005).
 - [GJ79] Michael R. Garey and David S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, 1979.
 - [GLL78] Ronald L. Graham, Shen Lin, and Chio-Shih Lin, *Spectra of numbers*, Mathematics Magazine **51** (1978), no. 3, 174–176.
 - [KNA94] Naoyosho Kanamaru, Takao Nishizeki, and Tetsuo Asano, *Efficient enumeration of grid points in a convex polygon and its application to integer programming*, International Journal of Computational Geometry and Applications **4** (1994), 69–85.
 - [Noo81] P. Noordzij, *On $(na) + (nb) = (nc) + (nd)$ for all positive integers n* , Nieuw Archief voor Wiskunde, III. Ser. **29** (1981), 59–70.
 - [O'B02] Kevin O'Bryant, *A Generating Function Technique for Beatty Sequences and Other Step Sequences*, Journal of Number Theory **94** (2002), no. 2, 299–319.
 - [PS90] Horacio Porta and Kenneth B. Stolarsky, *Half-silvered mirrors and Wythoff's game*, Canadian Mathematical Bulletin **33** (1990), 119–125.
 - [Ree57] John E. Reeve, *On the volume of lattice polyhedra*, Proceedings of the London Mathematical Society, II. Ser. **7** (1957), 378–395.
 - [Rez06] Bruce Reznick, *Clean Lattice Tetrahedra*, [arXiv.org:math/0606227](http://arXiv.org/math/0606227) (2006).
 - [Sca85] Herbert E. Scarf, *Integral polyhedra in three space*, Mathematics of Operations Research **10** (1985), 403–438.

- [Sto76] Kenneth B. Stolarsky, *Beatty sequences, continued fractions, and certain shift operators*, Canadian Mathematical Bulletin **19** (1976), 473–482.
- [Whi64] George K. White, *Lattice tetrahedra*, Canadian Journal of Mathematics **16** (1964), 389–396.